DNA REPLICATION AND TRANSCRIPTION IN THE *C. ELEGANS* EMBRYO

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

James Bellush

May 2019

DNA REPLICATION AND TRANSCRIPTION IN THE *C. ELEGANS* EMBRYO

James Bellush, Ph.D.

Cornell University 2019

Animal embryogenesis represents a cascade of genetically encoded events that transform a single-celled zygote, composed of cytoplasm and nuclei from two haploid gametes, into a fully formed organism. The “totipotent” zygote is the only cell with the ability to divide and produce cells of all differentiated types; however, the manner in which the embryonic genome is replicated, packaged, and transcribed during the regulatory transitions of animal development remain elusive.

To further our understanding of early animal development, I developed integrative deep sequencing and bioinformatics approaches to profile the genomic landscape of embryogenesis in the nematode *Caenorhabditis elegans*. Comparison of DNA replication, nascent RNA transcription, and active histone modification profiles from synchronized populations of staged worm embryos revealed a striking model of genome organization, shaped by the combined influence of DNA replication origin specification and the developmental transcriptome. We find that the embryonic transcriptome and chromatin modification landscape are largely defined by DNA replication origins and the constitutive transcription of housekeeping genes arranged within operons. We observe a dramatic remodeling of gene expression relative to

DNA replication origins during the shift from the proliferation to the differentiation phase of embryogenesis—analysis of embryonic chromatin states and nascent RNA transcription patterns reveal that the morphogenesis gene expression program in differentiated cells is coupled to the *de novo* remodeling of histone modifications and a switch to pervasive RNA polymerase elongation.

Our data suggest a novel mechanism by which a small regulatory RNA pathway acting in the nucleus is involved in coordinating the switch between the germline/embryonic gene expression program and the somatic differentiation program. This pathway centers on an essential *C. elegans* Argonaute, CSR-1, and its associated 22-nt small RNAs that mediate CSR-1 interaction with the nascent embryonic transcriptome. CSR-1 is known to balance the level of maternal RNA transcripts that support early embryogenesis through selective slicing of endogenous mRNA transcripts during germline development. Our model proposes that CSR-1 activity from the adult germline is inherited in the early embryo and that its gradual titration from somatic cells during early embryonic cell divisions helps gradually transition the embryo from a program of rapid proliferation to one of tissue-specific morphogenesis. The work described herein reveals the influence of DNA replication and transcriptional regulation on the spatiotemporal dynamics of *C. elegans* embryogenesis and lends insight into the evolutionary forces which shape metazoan genome structure and function.

BIOGRAPHICAL SKETCH

James Bellush grew up about 60 miles from New York City, in rural Sussex County, New Jersey. A constant curiosity for the natural world and an affinity for the outdoors initially drew James to the sciences; his formal scientific training began during his undergraduate studies at Moravian College in Bethlehem, Pennsylvania. His initial research interests were related to tropical biology and evolution, which drew him to a study-abroad research program in Monteverde, Costa Rica in the summer of 2009. Under the guidance of Dr. Alan Masters, James studied the various counter-adaptations defining plant-insect coevolutionary relationships in the Costa Rican cloud forest. Capitalizing on this research experience and following a desire to learn the principles of molecular biology, James returned to Moravian to continue his studies of the essential aspects of nucleic acid biochemistry, molecular evolution, and epigenetics in *Drosophila,* under the guidance of advisors Dr. Chris Jones and Dr. Shari Dunham.

Commencing his Ph.D. studies at Weill Cornell in 2012, James joined the lab of Dr. Iestyn Whitehouse at Memorial Sloan Kettering with an aspiration to learn about how genome transactions, such as gene transcription and DNA replication, pattern animal development. Work published by the Whitehouse lab has contributed fundamental insights into chromatin structure, gene transcription, and chromatin assembly during DNA replication. In the thesis work James has completed during his graduate studies, he applies a variety of genomics assays to study the regulatory program guiding early embryogenesis in the nematode *C. elegans*.

ACKNOWLEDGMENTS

Since my childhood, I have had a fascination with the complexity of life in the natural world—to have had opportunity to train and work as a scientist is a truly astounding gift that I could not have imagined would be possible. I find that it is important to acknowledge the scientific community, past and present, which fosters this amazing atmosphere of discovery that has broadened my horizons and opened doors to exciting opportunities. Being able to join to this lineage of scientists, who have contributed their research to better understand our natural world, is a truly incredible honor.

My guides through this scientific journey have been my mentors, whom I would like to acknowledge for their encouragement and guidance: Dr. Alan Masters (Monteverde, Costa Rica), Dr. Chris Jones (Moravian College), Dr. Shari Dunham (Moravian College), and Dr. Iestyn Whitehouse (Memorial Sloan Kettering).

My support system has always been my friends and family. To my parents, Joel and Kathy, I am grateful for their constant love and encouragement. To my wife Jenny, I am grateful to have a partner and a friend. This chapter of our time together has been filled with its ebbs and flows, but the greatest joy has come from sharing the moments with you.

TABLE OF CONTENTS

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

Caenorhabditis elegans: Ideal model of metazoan development

Embryogenesis is a cascade of genetically encoded events that transform a single celled zygote, composed of cytoplasm and nuclei from two haploid gametes, into a fully formed organism. For generations, biologists have sought to characterize the heritable components that guide the transformation of the embryo (Boveri, 1889). A central question has been whether specialized cell and tissue types arise from a gradual series of choices—"epigenesis"—or whether development is simply an expansion of "preformed" tissues in the egg (Noble, 2015).

A fundamental shift in the paradigm of developmental biology came when a research fellow named Conrad Waddington proposed that phenotypic complexity might arise from a series of cell fate decisions (Waddington, 1957). To frame his concept, Waddington represented the developmental landscape as a ball rolling down a sloped valley, carved by a network of diverging canals; the canals become deeper as the ball travels down the slope, signifying the gradual stabilization of cell fate as development proceeds (Waddington, 1957). Eventually, the ball comes to rest at a low stable minimum that embodies the terminal, or differentiated, state of the cell, while the path it took represents its developmental lineage (Waddington, 1957). The expansion of Waddington's idea that cell fate is established by the interaction of functionally distinct genes provided a logical framework for molecular biologists to

explore gene-phenotype relationships during animal development (Noble, 2015; Waddington, 1956).

This integrated approach provided a valuable starting point for the molecular biologist Sydney Brenner, who sought to develop an ideal model organism system that would illuminate the genetic pathways involved in animal development (Brenner 1974). Brenner's ultimate choice was the nematode *Caenorhabditis elegans*, a free-living, non-parasitic worm that measures 1-2 mm in length, roughly the size of a comma on a printed page (Figure 1.1) (Brenner, 1974). Fortuitously, the selection of this small worm not only provided an ideal system to study animal development, but also accelerated the discovery of previously unknown biological phenomena, such as apoptosis and RNA interference (Sulston & Horvitz, 1977; Fire et al. 1998).

Genetics and Development

The power of the *C. elegans* system lies in its experimental convenience and the platform it provides to study developmental genetics. In normal laboratory conditions, worms exhibit a short generation time, traversing embryogenesis, four intermediate larval stages, and reproductive maturity in approximately 3 days (Bird & Bird, 1991) (Figure 1.2). *C. elegans* populations mostly consist of hermaphrodite worms (XX) that possess a U-shaped gonad capable of generating both sperm and eggs; through self-fertilization, a single hermaphrodite can produce ~300 embryos in their lifespan, a feature that allows for the efficient isolation of homozygous recessive mutants (Brenner 1974; Villeneuve & Meyer 1987). *C. elegans* males (XO), which

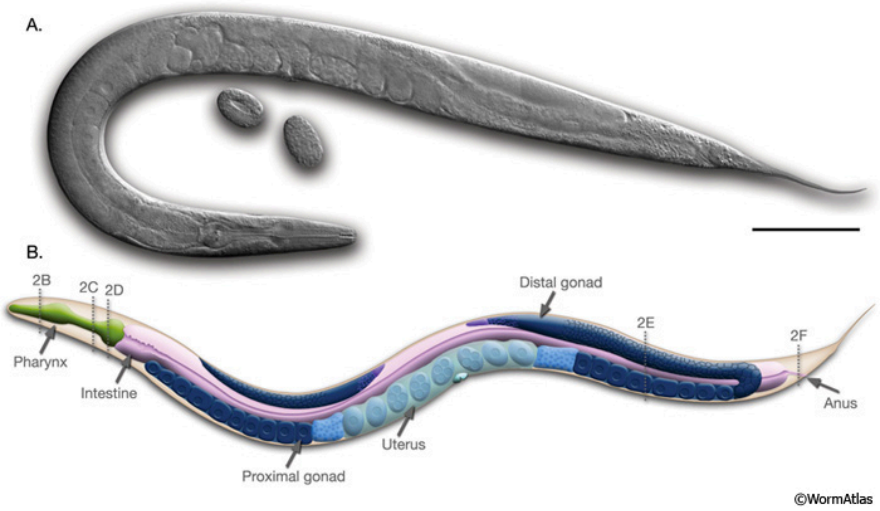

Figure 1.1. General anatomy of *Caenorhabditis elegans* adult hermaphrodite. A. DIC image of a gravid adult and embryos, left lateral side. B. Anatomical structures of adult, including pharynx (green), intenstine (pink), and the germline (blue). From Worm Atlas Database.

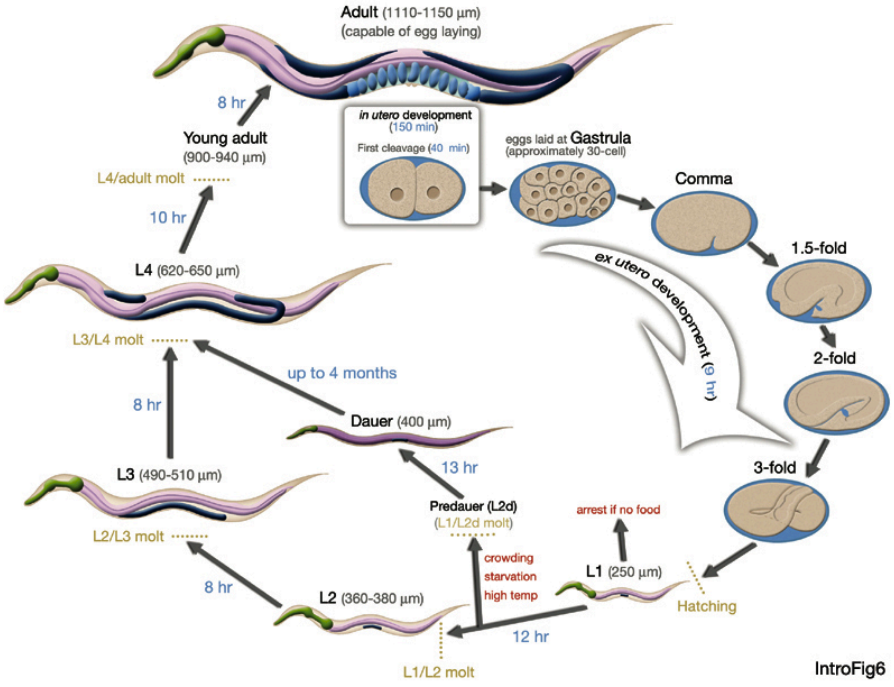

Figure 1.2. Life cycle of *C. elegans* at 22°C. Fertilization begins embryogenesis at 0 min; the first cleavage occurs 40 min post-fertilization; embryos are laid outside of the uterus at 150 min; blue numbers represent time spent at indicated larval stage. From Worm Atlas Database.

are anatomically and behaviorally distinct from hermaphrodites, also arise spontaneously at a low frequency in populations (Brenner et al. 1974). Males can be identified by a male-specific tail structure and distinct locomotion which is regulated by a set of sensory neurons that control male mating behavior (White et al. 1976, Liu and Sternberg 1995). Mating of males and hermaphrodites provides an opportunity to perform crosses among mutants to characterize genetic interactions among developmental pathways (Hodgkin et al. 2012).

Early in embryogenesis, germ cells are specified as distinct from somatic cells—the germline blastomeres, or P-blastomeres, acquire their potential from a master transcriptional repressor, PIE-1, and macromolecular complexes called P-granules (Figure 1.3A) (Strome and Wood 1983, Mello et al. 1996). The two primordial germ cells (PGCs) remain transcriptionally quiescent during embryogenesis and initiate proliferation of the germline in response to nutritional signals during the first larval phase (Figure 1.3B) (Crittenden et al. 1994; Henderson et al. 1994). The hermaphrodite germline is syncytial and exhibits distal-proximal polarity, with mitotically dividing germ cells at the distalmost end and meiotic cells extending proximally (Hirsh et al. 1976; Kimble & White, 1981; Strome 1986). Oocytes are produced during the entire adult reproductive span, while sperm are generated during the L4 stage, stored in a hermaphrodite specific structure called the spermatheca, and fertilize oocytes during adulthood (Ward and Carrel 1979, Shakes and Ward 1989).

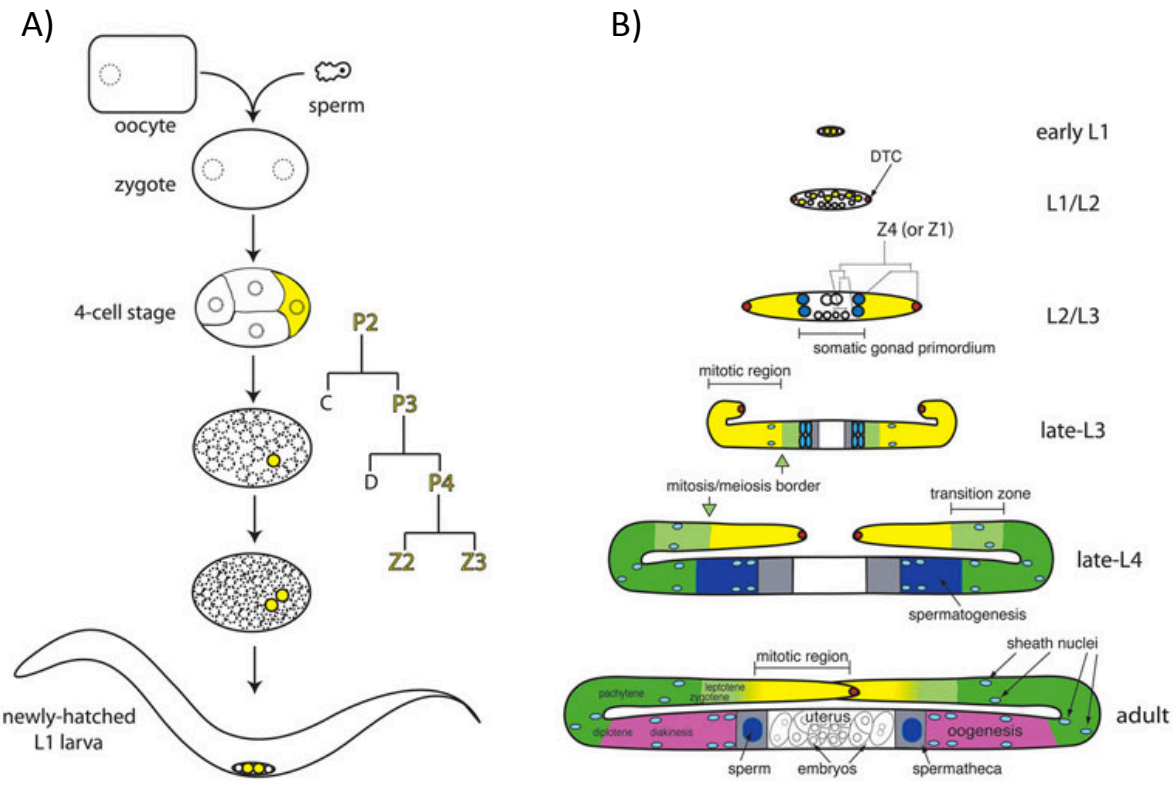

Figure 1.3. Germline development in *C. elegans*. A) The germline lineage (or P-lineage), depicted in yellow, is specified shortly after fertilization in the early embryo. The P-cells undergo 4 cell divisions before the primordial germ cells, Z2 and Z3, arrest at the 100-cell stage for the remainder of embryogenesis. B) Hermaphrodite post-embryonic germline development. Mitotic germ cells (yellow) proliferate synchronously during the L1-L2 stage before establishing a mitotic-meiotic cell gradient (yellow-to-green) in late-L3. The gradient depends on proximity to Notch mediated signaling by the distal tip cell (red) and post-transcriptional gene regulation by RNA binding proteins. Gametogenesis spans the period from late-L4 to adulthood– spermatogenesis produces around 300 sperm in the spermatheca (blue) at the L4 stage, followed by oogenesis (pink) throughout remaining adulthood. From Hubbard and Greenstein 2005.

Worms remain transparent for their entire life-cycle, providing the opportunity to examine developmental deviations non-invasively in the living animal (Sulston and Horvitz, 1977). Embryogenesis occurs within 14 hours at room temperature and takes place within a protective eggshell 50 μm in length (Wharton 1980). Microscopic observations of embryonic and post-embryonic development revealed that the *C. elegans* cell lineage is largely invariant; at the time of hatching, animals possess 558 somatic cells whose individual fate is determined by the embryonic lineage (Sulston & Horvitz, 1977; Sulston et al. 1983). The non-invasive screening platform and invariant cell lineage afforded by the *C. elegans* system accelerated the physical mapping and eventual sequencing of the first animal genome (Coulson et al. 1986; Cutter et al. 2009).

Genome Structure

The *C. elegans* genome is encoded by five autosomes and one sex chromosome; interestingly, the chromosomes are holocentric in nature, defined by many mitotic centromeres which assemble along the length of the chromosome to mediate chromosome segregation (Albertson and Thomson 1982; Coulson et al. 1986). Each chromosome has a central gene-dense region with discrete boundaries, flanked by gene-poor arm regions; comparison of recombination frequencies along chromosomes revealed that recombination events typically occur within the terminal arm regions at a frequency of one event per chromosome per meiosis (Hodgkin 1993, Pilgrim 1993, Barnes et al. 1995).

Sequencing and annotation of the worm genome revealed approximately 21,000 protein-coding genes, typified by short introns, minimal 5' and 3' untranslated regions (UTRs), and a clustered arrangement in chromosome centers (Saito et al. 2013; Blumenthal & Spietch, 1996; Spieth et al. 2014). However, characterization of the complete *C. elegans* transcriptome remains unfinished, as many gene promoters, transcription start sites (TSS), and non-coding RNA transcripts have not been completely annotated (Chen et al. 2014). Annotation of the transcriptome in *C. elegans* is complicated by the fact that a majority of protein-coding transcripts are modified by trans-splicing; this process involves the addition of a short 22 nucleotide (nt) RNA leader to the 5' end of the pre-mRNA, removing the true transcription start site (TSS) (Spieth et al. 2005). The process of trans-splicing is conserved among nematodes, however elucidating a specific role for trans-splicing in *C. elegans* development is an active area of inquiry (Spieth et al. 2014).

Life Cycle and Developmental Stages

The *C. elegans* natural habitat comprises any microbe-rich habitat with decaying plant material, such as compost heaps or rotting fruit (Cutter et al. 2009). A reliance of ephemeral food sources during the natural history of *C. elegans* led to the development of two distinct life-cycles; a rapid, reproductive life-cycle in the presence of food and a starvation, stress-induced cycle marked by distinct periods of developmental arrest (Felix and Braendle, 2010) (Figure 1.2). Well-fed hermaphrodite worms complete their life-cycle in 3-4 days marked by the passage of

four intermediate larval stages (L1-L4) and the renewal of egg laying once reaching mature adulthood (Ambros and Horvtiz, 1984, Wood et al. 1988). Animals that experience a period of starvation, crowding, or thermal stress enter a state developmental arrest known as diapause, which involves a general reduction of metabolism and an increase in stress resistance (Frezal and Felix 2015).

Animals can only enter diapause in the first two larval stages; embryos that hatch as L1 larva in the absence of food can diapause for up to 3 weeks, while starvation and crowding among L1-L2 larvae induce an alternative L3 morph known as the dauer (Baugh 2013, Elling et al. 2007). Dauer larvae are stress resistant for several months and are highly represented in wild populations—dauers develop a thicker cuticle and induce behavioral and metabolic pathways designed to facilitate the search of new food sources and colony formation (Ambros and Horvitz, 1984, Euling and Ambros 1996, Barriere and Felix 2007).

Studies of the genes that activate dauer formation (daf) in *C. elegans* found a surprising association between germline signaling and metabolism in long-lived worms (Hsin and Kenyon 1999, Gerisch et al. 2001). Activation of the dauer transcriptional cascade in animals that had their germlines removed extended life-span (Gerisch et al. 2001, Berman and Kenyon, 2006). This type of regulatory structure appears to be required to divert energy into somatic cell survival or high-cost germ cell reproduction depending on the environmental conditions (Finch 1990, Kenyon et al. 1993, Larsen et al. 1995).

In summary, studies to date suggest that the natural history of *Caenorhabditis elegans* has profoundly shaped the organization of the worm genome and the gene regulatory mechanisms that control cell developmental responses to intrinsic and extrinsic signaling.

The following section will briefly highlight the regulation of cell division, zygotic transcription, and lineage specification during early embryo development. In order to provide a broader evolutionary perspective on the developmental strategies among metazoan model organisms, this section cites examples of developmental events from *Drosophila, Xenopus*, mouse, and zebrafish, in addition to the specific examples from *C. elegans* embryogenesis.

Rapid Cell Divisions of Early Metazoan Embryogenesis

The "totipotent" zygote is the only cell with the ability to divide and produce cells of all differentiated types (Seydoux and Braun 2006). The molecular factors encoding this totipotency are cytoplasmic and nuclear in origin—maternal factors inherited from the egg cytoplasm encode factors that direct early embryonic cell divisions (e.g. histones, DNA replication factors, cell cycle kinases) and master regulators (e.g. transcription factors, coactivators, repressors) that initiate transcription of the zygotic genome (Draper et al. 1996, Liang et al. 2008). The regulatory interactions between embryonic programs of cell proliferation and differentiation coordinately controls establishment of cell fate, lineage-restricted gene expression, and tissue differentiation (Giraldez 2010; Lee et al. 2014).

In metazoans such as *Drosophila*, *Xenopus,* and *C. elegans*, early embryonic cell divisions are characterized by short cell cycles consisting of rapid DNA replication and no gap phases (Murray and Kirschner 1989; Edgar et al. 1994; Farrell and O'Farrell 2014). This adaptation likely ensures that eggs laid in a hostile, external environment proceed to hatching as quickly as possible (Farrell & O'Farrell 2014). During these early division cycles, DNA replication, chromosome segregation, and cell cycle control are governed by the store of maternally deposited gene products from the egg cytoplasm (Murray and Kirschner 1989, Edgar et al. 1994; Almouzni and Wolffe, 1995). In *Drosophila* and *Xenopus*, the first 10-14 nuclear divisions, referred to as cleavage cycles, are rapid, synchronous, and occur in the absence of zygotic transcription (Foe & Alberts, 1983). The high concentration of maternally provided replication factors leads to the activation of many closely spaced replication origins, which dramatically shortens S-phase to a length of 5-8 minutes (Kane and Kimmel 1993; Blumenthal et al. 1974). Measurement of DNA replication activity in 2-cell stage *Xenopus* embryos revealed that DNA replication origins are spaced every 10-15 kb; during the cell divisions of differentiated somatic cells, the inter-origin distance increases 10-fold (Hyrien et al. 1995; Cayrou et al. 2011).

Drosophila and *Xenopus* embryos form a hollow ball of cells during the cleavage cycles called a blastula—as more cells occupy the same volume within the embryonic space, the nuclear-cytoplasmic ratio increases and causes the titration of limiting replication factors (Edgar et al. 1986). The synchronous cell divisions in these organisms become markedly slower due to an increase in S-phase length,

caused by a reduction in the number of replication initiation events, rather than changes in the rate of the replication fork (Blumenthal et al. 1974). This dramatic change in cell cycle regulation in *Drosophila* and *Xenopus* occurs at a stage called the mid-blastula transition (MBT), which happens to coincide with the activation of zygotic transcription (Edgar & O'Farrell 1989, Amodeo et al. 2015; Collart et al. 2013).

Characterizations of early *C. elegans* embryonic divisions have revealed key developmental differences from the *Drosophila* and *Xenopus* programs (Figure 1.4) (Edgar & McGhee, 1988; Bao et al. 2008). Although the spacing of DNA replication origins has not been measured in *C. elegans*, analysis of cell cycle timing revealed that DNA replication is completed within 10-25 minutes (Edgar & McGhee, 1988). Unlike *Drosophila* and *Xenopus* embryos, the early divisions of *C. elegans* embryonic cells are asymmetric and asynchronous; early cell divisions proceed at a similar pace in the pre-gastrula embryo, however as development proceeds, each cell lineage adopts a unique cell division timing (Bao et al. 2008). Cell cycle pace appears stable throughout the proliferative phase of embryogenesis and does not exhibit a marked increase in S-phase length (Bao et al. 2008; Edgar & McGhee et al. 1988). While the specific mechanisms regulating DNA replication during early *C. elegans* embryogenesis, the distinctions from the other externally developed metazoan embryos may be related to the early fate specification of the *C. elegans* cell lineage (Mello et al. 1992).

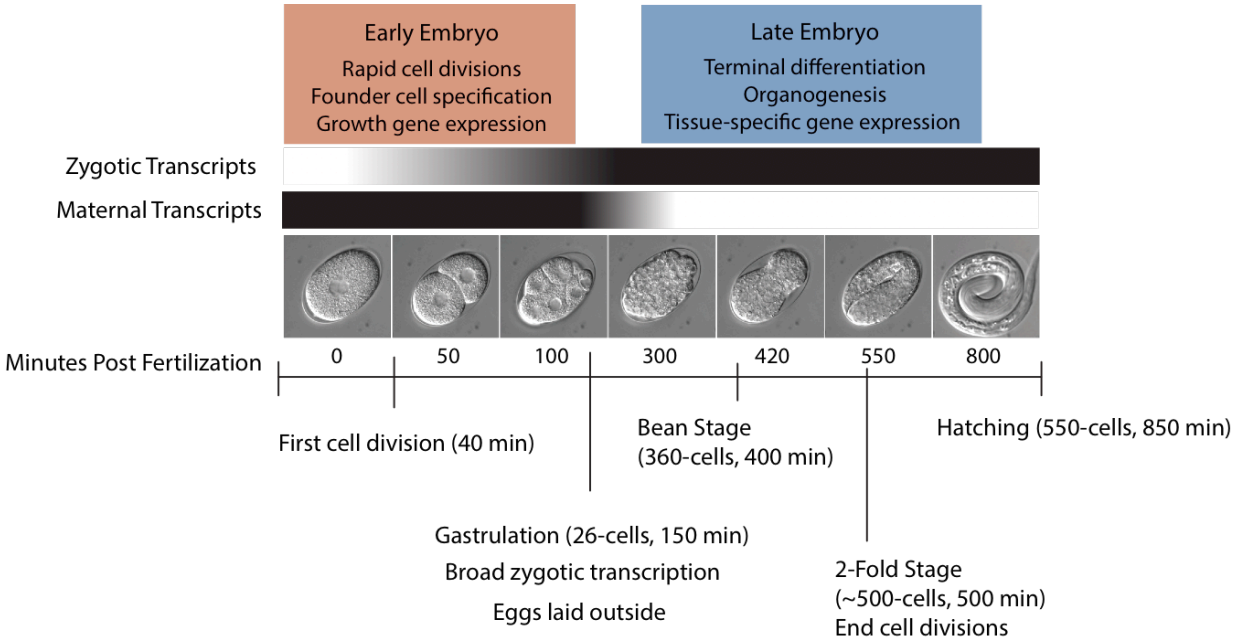

Figure 1.4. Timeline of *C. elegans* embryogenesis . Early embryonic divisions begin ~40 minutes after fertilization and are driven by maternal gene products encoding cell cycle factors. Gastrulation coincides with the beginning of zygotic transcription and ex utero development; the zygotic transcriptome gradually transitions from away from proliferation and towards somatic differentiation programs. The final cell divisions are complete by 500 minutes and are followed by terminal differentiation, organogenesis, and hatching of the 550-cell animal.

Cell Fate Specification in *C. elegans* Embryos

The invariant cell lineage of *C. elegans* relies on the specification of the "founder cells", AB, MS, E, C, and D—the consistent cell-cell interactions set up by reproducible division patterns of the founder cells are essential for the spatial organization of the developing embryo (Figure 1.5A) (Schierrenberg 2016). Most importantly, the founder cell divisions asymmetrically segregate developmental potential to each lineage in the form of master transcription factors and RNA binding proteins (Hutter & Schnabel 1994, Mango et al. 1994, Mello et al. 1994). The analysis of mutants displaying lineage defects revealed that the segregation of developmental potential to each founder cell relies on post-transcriptional and translational control of the maternal RNA transcripts encoding the specification factors (Mello et al. 1992).

The premature translation of a transcript or the misrouting of RNA binding proteins to the wrong founder cell causes a myriad of developmental defects that are a consequence of cell fate reprogramming (Horvitz & Sulston, 1980). The regulation of this process is required immediately following fertilization, beginning with the establishment of anterior-posterior cells of the embryo in the first cell divisions (Figure 1.5B) (Mello et al. 1992; Draper et al. 1996). In the 2-cell embryo, the segregation of the MEX-3 RNA binding protein to the posterior cell specifies the production of numerous muscles in the worm, while the anterior cell lacks this potential; mutations in *mex-3* result in an embryonic lethal phenotype, characterized by the duplication of the posterior muscle patterns in the anterior daughter cell lineage (Draper et al.

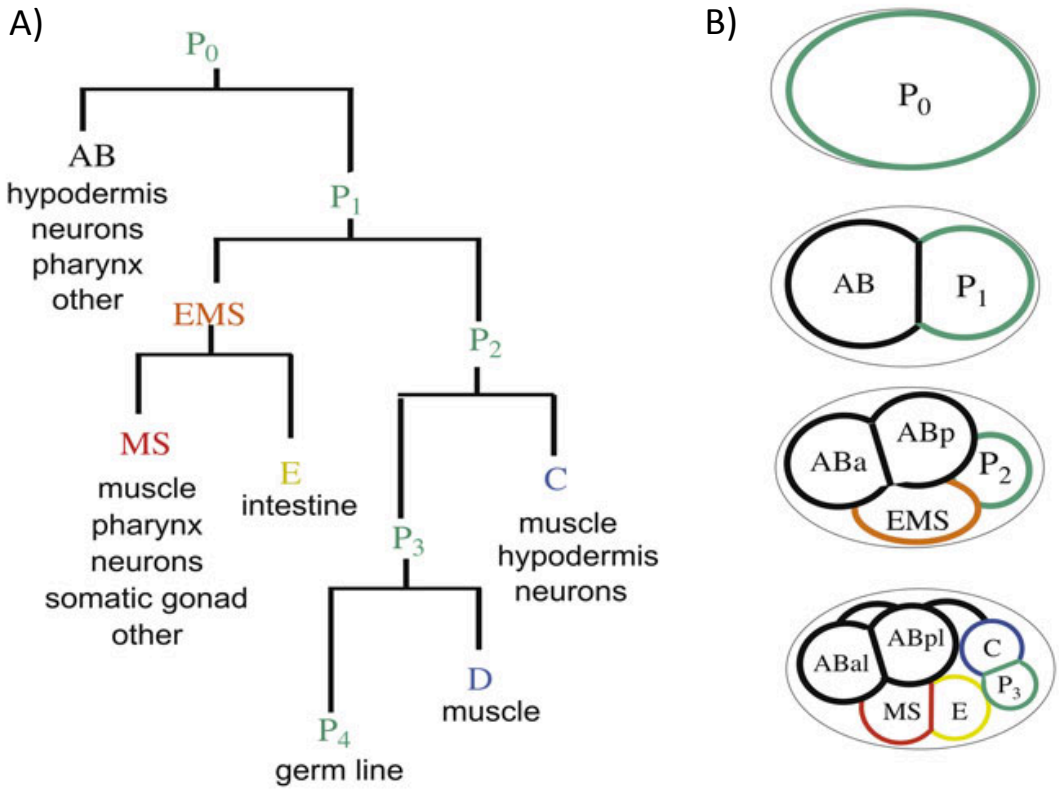

Figure 1.5. Asymmetric divisions establish *C. elegans* cell lineages during embryogenesis. A) Dendrogram of the early cell lineage after fertilization (P_0). Founder cells generated by asymmetric divisions are indicated with a different color, while cell cycle timing is represented by the length of each branch in the lineage. The daughter cell types of each founder cell are listed below. B) Spatial organization of founder cells. After the establishment of the anterior-posterior axis (left and right in schematic, respectively) is established by fertilization, founder cell positioning is regulated by actomyosin networks, polarity proteins, and polarity mediators that segregate cytoplasmic components to the appropriate daughter cell. From Gonzcy and Rose 2005.

1986). A similar lineage duplication is observed in mutants for the bZIP transcription factor SKN-1, which is essential for specifying the intestinal and muscle cells of the E and MS lineages, respectively (Bowerman et al. 1992). Loss of function mutants for *skn-1* fail to activate endoderm-specific gene expression, leading to the reversion of the E lineage to the MS lineage, characterized by mutant animals with extra hypodermal and muscle cells in place of the intestinal cells (Bowerman et al. 1992).

Decades of analyzing *C. elegans* lineage mutants has identified many of the regulatory factors controlling cell fate specification in the embryonic and post-embryonic lineages (Schierrenberg, 2016). The unique "preformistic" mode of embryogenesis deployed by *C. elegans* depends on the unequal distribution of cell fate through asymmetric divisions (Kelly 2014). Quantitative analysis of lineage specific transcriptomes by single-cell RNA sequencing have confirmed findings from the genetic analysis of lineage mutants, concluding that many of the differentially segregated transcripts among founder cells encode transcription factors that activate the expression of genes in somatic cell lineages (Hashimony et al. 2012). However, the influence of cell fate specification on DNA replication and gene transcription patterns in the early *C. elegans* embryo remains unknown and will be explored in Chapter 2 of this work.

The Maternal-to-Zygotic Transition

In most metazoans, maternal gene products are essential for oocyte maturation and driving early embryonic cell divisions—but eventually the maternal

instructions become incompatible with the later developmental events (Walser & Lipshitz 2011). In a process known as “maternal clearance”, the RNAs and proteins from the egg cytoplasm are deleted and zygotic regulatory instructions are installed in their place through the activation of zygotic gene transcription, commonly called zygotic genome activation (ZGA) (Stoeckius et al. 2014).

As was mentioned above, *Drosophila* and *Xenopus* embryos carry out the first 10-14 cycles in the absence of zygotic transcription; while the regulatory logic of ZGA is an area of active investigation, current models consider the interactions between maternal transcription factors and zygotic gene regulatory sites as critical to the process (Lee et al. 2014). One model proposes that maternal repressors initially block assembly of transcription complexes at zygotic gene promoters but are titrated away as the nuclear to cytoplasm ratio increases, allowing zygotic gene transcription to initiate (Newport & Kirschner 1982). A “maternal clock” model has also been proposed, in which the increase in activity or quantity of a maternal factor triggers transcription after a threshold level has been reached—this model is supported by the complex network of factors which regulate maternal RNA translation in oocytes and early embryos (Richter & Lasko 2011).

The requirement to slow the cell cycle prior to zygotic genome activation may be related to the time necessary to transcribe and process longer genes, which otherwise would be interrupted by passing replication forks—indeed, the earliest expressed genes in *Xenopus* and *Drosophila* are short and oriented co-directionally with DNA replication forks (Rothe et al. 1992, Lee et al. 2014). In *Drosophila,*

Xenopus, and mouse, the first wave of zygotically transcribed genes encode basic cellular functions, such as DNA replication, transcription, and RNA processing, while the second wave encodes the master regulatory factors that initiate a subsequent cascade of gene expression programs in differentiated cells (Lee et al. 2014; Hamatani et al. 2004). Metazoan species show extensive variability in the developmental timing of ZGA, reflecting that animal specific regulatory contexts influence the transcriptional competency of the zygotic genome (Hug et al. 2017, Ke et al. 2017).

Differential gene expression in the embryonic cells depends on the accessibility of regulatory sites to transcription factors and RNA polymerase (Chen et al. 2011). Chromatin represents the macromolecular complex of genomic DNA and protein that packages the DNA within the nucleus, but also regulates its biochemical activities (Kornberg 1974; Grunstein 1997). During ZGA, the accessibility of regulatory sequences within chromatin influences which genes are transcribed into RNA (Perino & Veenstra 2016). The following section will describe the principal mechanisms of gene regulation by chromatin, including establishment of promoter architecture, modification by post-translational marks (PTMs), and formation of topological domains.

Gene Regulation and the Transcription Cycle

The logic of gene regulation was first proposed by Jacob and Monod in the 1960s—they postulated that *cis*-acting DNA sequences at the transcription start sites of genes control transcription activity by recruiting diffusible gene products in *trans*

(Jacob & Monod 1960). Indeed, regulatory proteins we now call transcription factors (TFs) compete for interaction with cis-regulatory elements, often housed in clusters called promoters or enhancers, to promote or repress gene transcription (Brent & Ptashne, 1984).

The principles and mechanisms underlying transcription of a DNA template are remarkably conserved across the three domains of life (Figure 1.6) (Hahn 2004). The typical RNA polymerase II transcription cycle begins with the binding of TFs to core promoter sequences upstream of the transcription start site (TSS) (Figure 1.6A) (Dynan and Tijan, 1983; Ptashne and Gann, 1997). Coactivator complexes such as Spt/Ada/Gcn5 acetyltransferase (SAGA) or Mediator in eukaryotes, facilitate the binding of the general transcription factors (GTFs) that position RNA polymerase at the promoter (Figure 1.6B) (Grant et al. 1997, Dynlacht et al. 1991, Verrijzer and Tijan, 1996). The GTFs stimulate the formation of the Pol II preinitiation complex (PIC), featuring a sequential addition of the transcription factors TFIID, TFIIA, TFII B, RNA polymerase II, and TFIIE that stabilizes a paused transcription complex in a closed conformation proximal to the TSS (Figure 1.6C) (Buratowski et al. 1989, Roeder 1991).

The transition from the closed PIC to the active open PIC involves ATP hydrolysis to melt the DNA duplex, synthesis of the RNA transcript, and migration towards the TSS (Holstege et al. 1996, Murakami et al. 2002). The switch to productive elongation by RNA Pol II is regulated through protein-protein interactions on the C-terminal domain of RNA polymerase, which is differentially phosphorylated

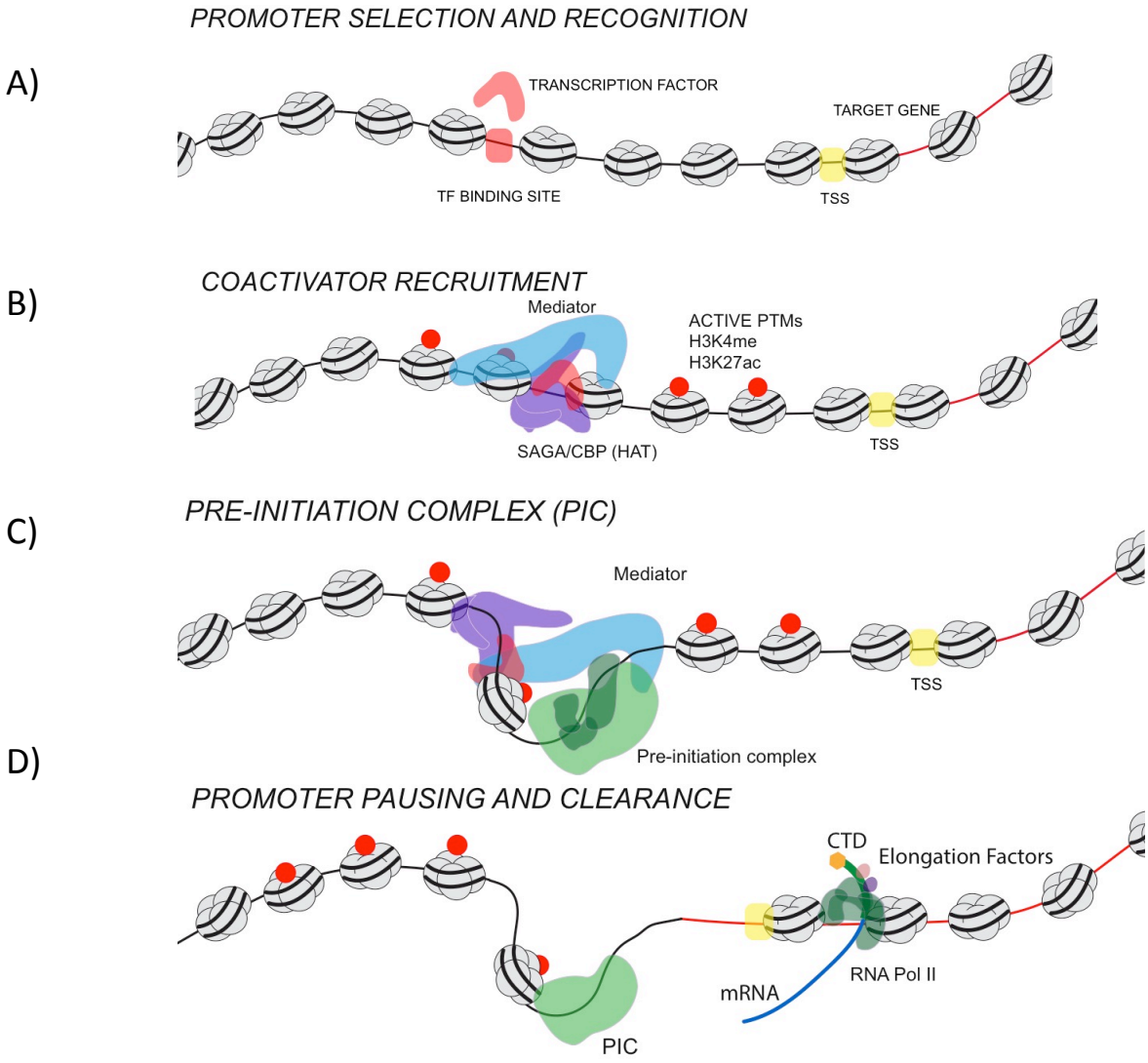

Figure 1.6. Transcription cycle of RNA Polymerase II regulated at gene promoters. A) Upstream activating sequences in the promoter interact with transcription factors. B) Coactivator complexes like Mediator and CBP locally remodel the promoter structure to increase the binding affinity of the general transcription complex. C) Pre-initiation complex (PIC) assembly stabilizes RNA Pol II at the promoter. D) Productive elongation requires numerous nuclear factors to associate with RNA Pol II to coordinate mRNA transcript processing.

during the transcription cycle to regulate RNA transcript processing (Figure 1.6D) (Goodrich and Tijan 1994, Krumm et al. 1995, Mandal et al. 2004).

C-Terminal Domain of RNA Polymerase II

While RNA Pol II is transcribing, an exceptional diversity of nuclear processes must be coordinated, including RNA modification, splicing, chromatin modification, and RNA processing—the major scaffold through which these pathways engage with active transcription is the largest subunit of RNA polymerase, the C-terminal domain (CTD) (Phatnani and Greenleaf 2006). The CTD is unique in design—it is inherently unstructured and contains a conserved heptad repeat sequence (YSPTSPS) represented in 25-52 tandem copies, depending on the species (Corden 1990). Hyperphosphorylation of the CTD at Ser2 and Ser5 are correlated with biochemical changes in Pol II holoenzyme structure and activity (Weeks et al. 1993). The distinct phospho-CTD states correspond to the physical association of nuclear factors to Pol II at distinct stages of the transcription cycle, providing a spatiotemporal link between pre-mRNA processing and elongating RNA polymerase (Phatnani and Greenleaf 2006).

CTD phosphorylation recruits factors required for 5' end capping, 3' end cleavage, and polyadenylation of the nascent RNA transcript (Rasmussen and Lis 1993, Proudfoot 2004). The localization of these processing factors to the appropriate region of the gene is also encoded by the CTD phosphorylation status; the phosphorylation of Ser5 by the Kin28/CDK7 kinase occurs at the 5' ends of

genes and recruits RNA capping enzyme, whereas the kinase CTDK-I/CDK9 phosphorylates Ser2 on the CTD, which recruits 3' end formation factors that carry out cleavage and polyadenylation of the transcript (Ho and Shuman 1999, Lindstrom and Hartzog 2001). The CTD of RNA Polymerase II facilitates the assembly of distinct regulatory complexes that alter transcriptional behavior according to spatiotemporal signals or cell specific regulation.

Architecture of Gene Promoters

Transcription by Pol II is regulated by chromatin structure at gene promoters, where transcription factors are competing for access to cis-regulatory sequences (Levine and Manley, 1989). The promoter architecture can be modulated by conserved families of chromatin interacting proteins that control gene activity by altering the density, positioning, and composition of nucleosomes (Mizuguchi et al. 1997, Tsukiyama & Wu, 1995, Wysocka et al. 2006).

Nucleosomes are composed of two copies of each core histone H2A, H2B, H3, and H4; the wrapping of 147 bp of duplex DNA around each nucleosome constitutes the basic repeating unit of chromatin (Luger et al. 1997). Studies in budding in yeast have provided valuable insight into the diversity of promoter structures and how configurations of nucleosomes correlate to gene activity (Almer et al. 1986, Gregory et al. 1998). For example, promoter sequences help to create the optimal nucleosome positioning to buffer the binding dynamics of transcription factors—regulatory sites are often clustered in A/T rich sequences called nucleosome

depleted regions (NDRs) that remain accessible for TF binding at constitutively expressed genes (Segal et al. 2006, Yuan et al. 2005, Lee et al. 2004).

Coding and Non-Coding Transcription from Promoters and Enhancers

The concentration of coactivators and RNA polymerase at constitutive gene promoters has been associated with bidirectional transcription activity in most eukaryotes (Xu et al. 2009, Kim et al. 2010, Adachi and Lieber, 2002). Bidirectional promoters are often constitutively transcribed and associated with high levels of histone acetylation and transcription factor binding sites (Neil et al. 2009, Scruggs et al. 2015). Pervasive transcription has also been shown to be a conserved mechanism of generating regulatory non-coding RNAs in animals; examples include miRNA generation *Drosophila,* piRNA precursor generation in *C. elegans*, and endogenous siRNAs in mouse oocytes (Okamura et al. 2008, Gu et al. 2012, Watanabe et al. 2008). Analysis of nascent RNA sequencing data from mammalian cells suggests that, even in the absence of non-coding transcript function, the pausing and elongation cycles of RNA polymerase at these pervasively transcribed promoters reinforces an architecture to promote continuous assembly of active transcription complexes (Core et al. 2008, Core et al. 2014). Identification of non-coding RNA functions during animal development provide interesting insights into how transcriptional activity shapes the evolution of animal genome structure (Stefani and Slack 2008).

Annotation of gene enhancers has also revealed evidence of constitutive transcription activity (Kim et al. 2010, Wang et al. 2011). Functional studies suggest that increasing the local concentration of RNA Pol II and its cofactors at enhancers increases the likelihood of transcript elongation and expression (Gerster et al. 1986, Weber and Schaffner, 1985). Limiting regulatory proteins, such as conserved transcription elongation factors SPT5 and P-TEFb, stimulate release of Pol II from the promoter into the gene body, a step in the transcription cycle referred to as promoter escape (Fuda et al. 2009). The pause time of RNA Pol II at enhancers is substantially shorter than it is at promoters, possibly reflecting a greater local concentration of pause-release factors (Jonkers and Lis, 2015, Henriques et al. 2018). Pervasive transcription of enhancer sites has also been suggested to concentrate transcription factors and RNA polymerase in proximity to target genes, even when the genes are not actually expressed (De Santa et al. 2010, Sigova et al. 2015). In sum, this evidence suggests that maintaining sites of pervasive transcription are important for regulating the developmental transcriptome and in turn, the overall genome structure (Levine and Tijan, 2003, Kvon et al. 2014).

Chromatin Modifications Associated with Transcription

Inducible gene transcription in response to signaling pathways often is preceded by a local modification of the chromatin structure (Almouzni et al. 1994, Tsukiyama et al. 1994). The recruitment of chromatin interacting proteins to target gene promoters stabilizes basal transcription machinery to increase the likelihood

that productive elongation may occur (Zhong et al. 2008, Ogryzko et al. 1996). The metastable character of chromatin imparts tremendous developmental control of gene activity—reversible post-translational modifications (PTMs) are associated with the waves of transcription from promoters in response to metabolic, differentiation, or stress signals (Li et al. 2007). The most-often studied PTMs in terms of their associated enzymes and regulatory function are the acetylation and methylation of lysines on histone H3 (Grunstein 1997, Krogan et al. 2003).

The crystal structure of the nucleosome presents N-terminal histone tails that protrude from nucleosome core and make contact with adjacent nucleosomes, suggesting that the overall local chromatin structure could be altered through inter-nucleosomal interactions of modified histone tails (Luger et al. 1997, Bannister and Kouzarides 2011). In addition to local structural changes, PTMs recruit chromatin complexes possessing PTM recognition domains—these peptides include methyl-lysine binders like chromodomains, Tudor domains, MBT domains, and PHD fingers, as well as the acetyl-lysine associated bromodomain (Huang et al. 2006, Mujtaba et al. 2007).

Histone Acetylation

Histone acetylation is catalyzed by histone acetyl-transferase (HATs) enzymes which transfer an acetyl group from an acetyl-CoA donor to the lysines of histone tails (Bannister and Kouzarides 1996). The weakening of electrostatic interactions between DNA and histones attributable to lysine acetylation has implicated HAT

enzymes in transcriptional coactivator complexes (Wolffe and Pruss, 1996). It was initially proposed that histone acetylation is developmentally regulated and correlated with the induction of zygotic transcriptional activation (Almouzni et al. 1994). Indeed, the establishment of developmentally regulated hubs of transcription, called enhancers, has been directly linked to HAT enzymes and histone acetylation (Ogryzko et al. 1996). Enhancers are identified by their ability to activate transcription at promoters independent of distance or orientation (Ptashne 1986). Annotation of active enhancers in different cell types and tissues revealed the ubiquitous presence of the transcriptional coactivator and HAT p300 or Creb binding protein (CBP) (Chrivia et al. 1993, Hatzis and Talianidis 2002). The binding of CBP to gene promoters increases local histone acetylation, in turn recruiting coactivators to promoter nucleosomes through interactions between bromodomains and acetylated histone tails (Hassan et al. 2002).

Histone Methylation: COMPASS Complex

The analysis of *Drosophila* mutants involved in regulating homeotic gene expression revealed the existence of two pathways which had opposing effects on gene expression: the trithorax group (trx) and the Polycomb group (Pc) of proteins (Kuzin et al. 1994, Jones and Gelbart 1993).

The trx group was later shown to encode a family of proteins which function as a highly conserved H3K4 methyltransferase complex—in mammals, the complex is called MLL (mixed-lineage leukemia), named for its relation to the development of

hematological malignancies, while in yeast the complex is called COMPASS (Complex Proteins Associated with Set1) (Miller et al. 2001). The complex contains the catalytic methyltransferase, Set1, as well as several subunits which regulate complex assembly, RNA processing, and the pattern of H3K4 mono-, di-, and trimethylation (Miller et al. 2001, Krogan et al. 2003). The COMPASS complex cotranscriptionally interacts with RNA Pol II through the interaction with the CTD and has shown genetic interactions with many pathways involved in the transcription cycle (Qiu et al. 2009, Wood et al. 2003, Xiao et al. 2003). In yeast, the level of H3K4 methylation has been shown to vary between the promoter (mostly H3K4me3) and coding region of the gene (mostly H3K4me2/1) (Pokholok et al. 2005). At the 5' end of the gene, when RNA Pol II is transcribing slowly during promoter escape, Set1 has a longer window of opportunity to methylate H3K4—during elongation, when RNA Pol II moves faster, less time is spent with its substrate, explaining the transition to monomethylation at the 3' end of genes (Wood et al. 2007). The differential patterns of methylation were eventually shown to differentially recruit chromatin cofactors to either the promoter or the 5' end of genes, thus establishing optimal chromatin structure for Pol II initiation and elongation (Kim and Buratowski 2009). The distribution of H3K4me domains at the 5' end of genes has been linked to transcriptional output at metazoan genes as well, where the broadest H3K4me domains at 5' ends of genes correlated with transcriptional consistency (Benayoun et al. 2014). The extensive genetic interactions between the COMPASS complex and nuclear factors involved in the transcription cycle suggest the existence of more

relationships between H3K4 methylation patterns and the targeted recruitment of transcription regulatory factors.

Histone Methylation: Polycomb Group

The function of Polycomb Repressive Complex 2 (PRC2) is to mediate gene silencing, largely through the methylation of lysine 27 on histone H3 (H3K27me) (Margueron et al. 2009, Jiao and Liu 2015). H3K27me marked chromatin is associated with diverse transcriptional processes, including the stable propagation of gene expression states during early development, heterochromatin formation at centromeric repeat regions, and X-chromosome silencing (Soshnikova and Duboule, 2009, Jacob et al. 2009, Simon et al. 2013). Similar to its counterpart COMPASS complex, the PRC2 complex can adopt many cell and tissue specific conformations—the core PRC2 complex is composed of four subunits: Ezh1/2, Suz12, Eed, and RbAp46/48, while additional subunits of note include the Jumonji demethylase subunit Jarid2, and the Polycomb-like subunits Pcl1/2/3 (Margueron and Reinberg 2011).

PRC2 complexes are able to maintain H3K27 methylation in consistent chromatin domains through recognition of the histone mark by the EED-EZH2 complex (Cao et al. 2002). Genome-wide ChIP sequencing in human embryonic stem cells found the relative abundance of methylation on H3K27 to be 50% dimethylated, 15% trimethylated, and 15% monomethylated, supportive of a model for H3K27me homeostasis by PRC2 (Peters et al. 2003). A large number of genes

involved in somatic processes are repressed by PRC2; H3K27me is associated with the epigenetic state of stem cells and is associated with repression of somatic differentiation during early embryogenesis (Agger et al. 2007, Gaydos et al. 2014, Boyer et al. 2006). However gene expression in PRC2 regulated domains can be induced by the action of H3K27 demethylases such as the Jumonji C domain family of proteins, which was reported to associate with PRC2 to regulate the level of H3K27me (Trojer and Reinberg 2007, Swigut and Wysocka 2007). Continuing questions regarding PRC2 function focus on how PRC2 is directed to regulated promoters at appropriate developmental times and how the modulation of H3K27me levels by demethylases or acetylation of H3K27 influences gene activity (Jiao & Liu 2015).

The fact that functionally opposing, essential complexes trx and PRC2 were isolated in the same *Drosophila* screen for developmental mutants speaks to an important role for chromatin in balancing the distribution of transcriptional activity in a complex genomic landscape (Jones and Gelbert 1993). A similar antagonistic relationship between histone methyltransferase complexes maintains the balance of gene expression in the *C. elegans* germline (Holdeman et al. 1998, Korf et al. 1998, Fong et al. 2002). The following section will briefly review the roles of the MES proteins in regulating somatic and germline gene expression programs during the *C. elegans* germline cycle.

Transgenerational Histone Modifications in the *C. elegans* Germline

Regulation of gene expression at the transcriptional and post-transcriptional levels is crucial for the proper specification, proliferation, and differentiation of germ cells (Mello et al. 1996, Tabara et al. 1999, Eckmann et al. 2002). In the primordial germ cells of the early embryo, the P-granules and the transcriptional repressor PIE-1 maintain germ identity by repressing RNA Pol II elongation and regulating the translation of mRNAs that specify the germ lineage (Seydoux and Braun 2006). During each cycle of germline proliferation, the maternal-effect-sterile (MES) proteins buffer germline gene transcription through opposing chromatin modification pathways—the H3K36 methylation pathway, mediated by MES-4, and the H3K27 methylation pathway, maintained by the MES-2/3/6 complex (MES complex) (Bender et al. 2006, Yuzyuk et al. 2009).

The gene family was named for the essential nature of all four MES proteins for germ cell viability, as the absence of maternally provided MES product in the embryo leads to the death of nascent germ cells and sterile adults (Capowski et al. 1991, Xu et al. 2001). The MES complex is the worm orthologue of the conserved PRC2 complex that modifies and binds H3K27me3 histones in repressed chromatin while MES-4 is a homologue of the vertebrate nuclear receptor binding SET domain protein (NSD-1) and is responsible for directing H3K36 methylation in germline transcribed regions (Korf et al. 1998, Lucio-Eterovic et al. 2010, Bender et al. 2004). MES-4 expression and activity are limited to germ cells, while MES complex activity is widespread in all cells (Reichsteiner et al. 2010, Capowksi et al. 1991).

H3K36me3 and H3K27me3 exhibit mutually exclusive patterns on *C. elegans* chromosomes, leading to the suggestion that the antagonistic action of of these domains maintains appropriate gene activity in the germline (Gaydos et al. 2012, Patel et al. 2012). While the functional consequence of histone modifications on transcription is not yet characterized, RNAi knockdown of MES-4 resulted in the silencing of germline expressed genes coincident with a decrease in H3K36me3 (Gaydos et al. 2014). The silencing of germline gene expression was attributed the encroachment of H3K27me3 from MES complex silenced regions on the autosomes and the X-chromosome, which is transcriptionally silenced in the germline through the dosage compensation pathway (Gaydos et al. 2012, Chuang et al. 1994). Altogether, the current model of antagonistic MES-4 and the MES complex activity suggests they are important in maintaining highly stable domains of gene expression in the germ line (Kelly and Fire 1998, Garvin et al. 1998, Paulsen et al. 1995).

Exhaustive ChIP-sequencing efforts to characterize histone modifications across *C. elegans* development have revealed a surprisingly stable association with genomic features, including housekeeping genes, germline-specific genes, and somatic differentiation genes (Figure 1.7) (Gu and Fire 2010, Daugherty et al. 2017, Liu et al. 2011). The distribution of histone modifications between embryos, containing mostly undifferentiated cells, and L3 larvae, representing differentiated animals, is nearly identical (Evans et al. 2016). The rapid lifecycle, post-transcriptional regulation by RNAi, and gene organization of *C. elegans* may have favored a minimal chromatin regulatory system, whose role is to buffer against

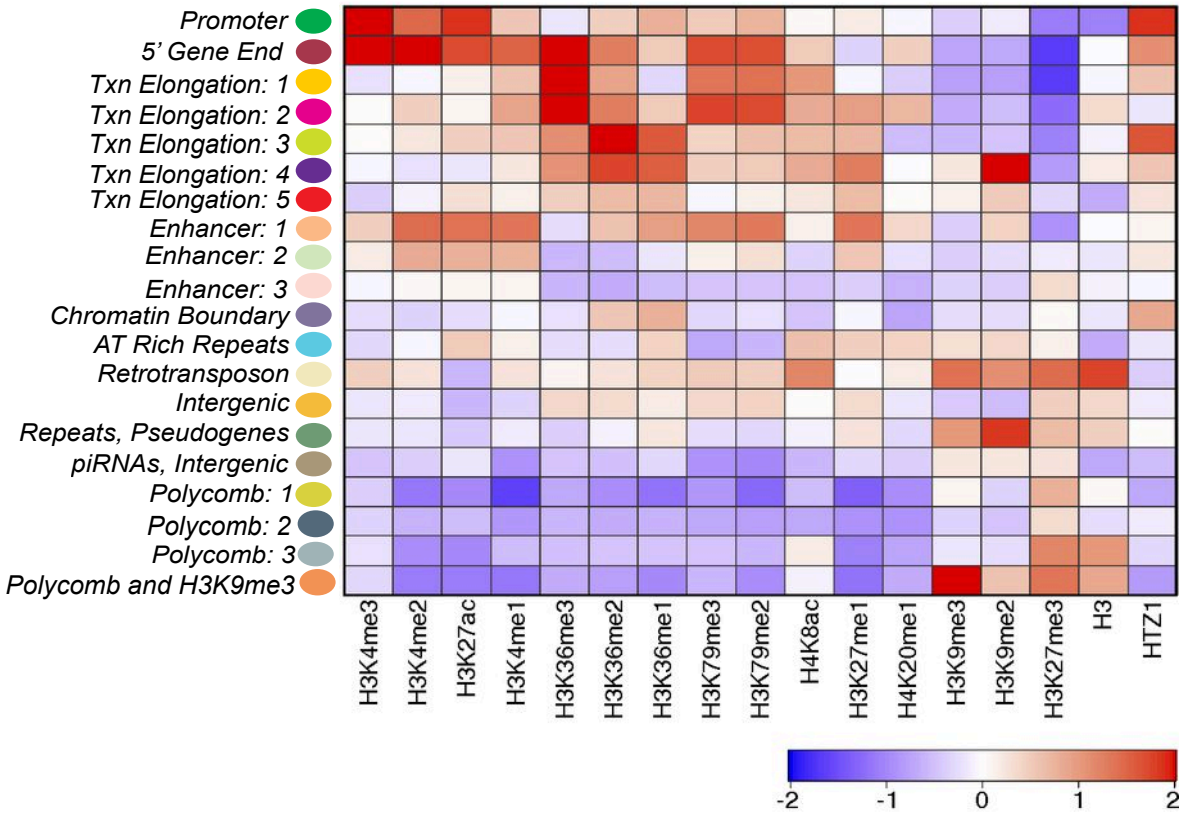

Figure 1.7. Chromatin states associated with *C. elegans* histone modifications. Adapted from Evans et al. 2016. Associations between annotated chromatin states (left axis, labeled by color) described by Hidden Markov Models (HMMs), and histone modifications (X-axis) were calculated at features in worm embryos and L3 larvae. Among the features included were transposons, highly transcribed promoters, germline-expressed transcripts, inducible, gene promoters, chromatin domain borders, and developmentally regulated genes. The log enrichment ratio (experimental/input) was calculated, normalized, and the scale bar shows the z-score of the mark averaged over 500 bp genomic windows. Chromatin states defined by these features were highly similar between embryos and L3 larva, supporting evidence of extensive maintenance of transcription and chromatin states across developmental stages.

changes to transcriptional states, rather than actively initiate transcriptome remodeling on a large chromosomal scale (Zaslaver et al. 2011, Buckley et al. 2012, Maxwell et al. 2012). However, characterizing how the DNA replication and transcription apparatus interact within these stable chromatin domains might be informative as to how the genomic information is expressed during developmental transitions, such as gametogenesis and early embryogenesis. This next section reviews the importance of gene organization during development—to illustrate these concepts, two distinct gene regulatory strategies will be discussed: topologically associating domains and operons.

Genome Organization: Topologically Associating Domains

Microscopic examination of chromosomes arranged within eukaryotic nuclei reveal discrete chromosome territories, suggesting that spatial organization is imparted within the nucleus (Manuelidis and Borden, 1988). A technological innovation called chromosome conformation capture, which could measure chromatin interactions by sequencing, eventually confirmed the existence of chromosome interaction domains called topologically associating domains, or TADs (Dekker et al. 2001, Lieberman-Aiden et al. 2009, Dixon et al. 2012).

The functional organization of the genome into topologically associating domains appears to relate to conserved replication and transcription domains during animal development (Pope et al. 2014, Sexton et al. 2012, Lupianez et al. 2015) (Figure 1.8). Strikingly, the replication timing profiles throughout S-phase revealed

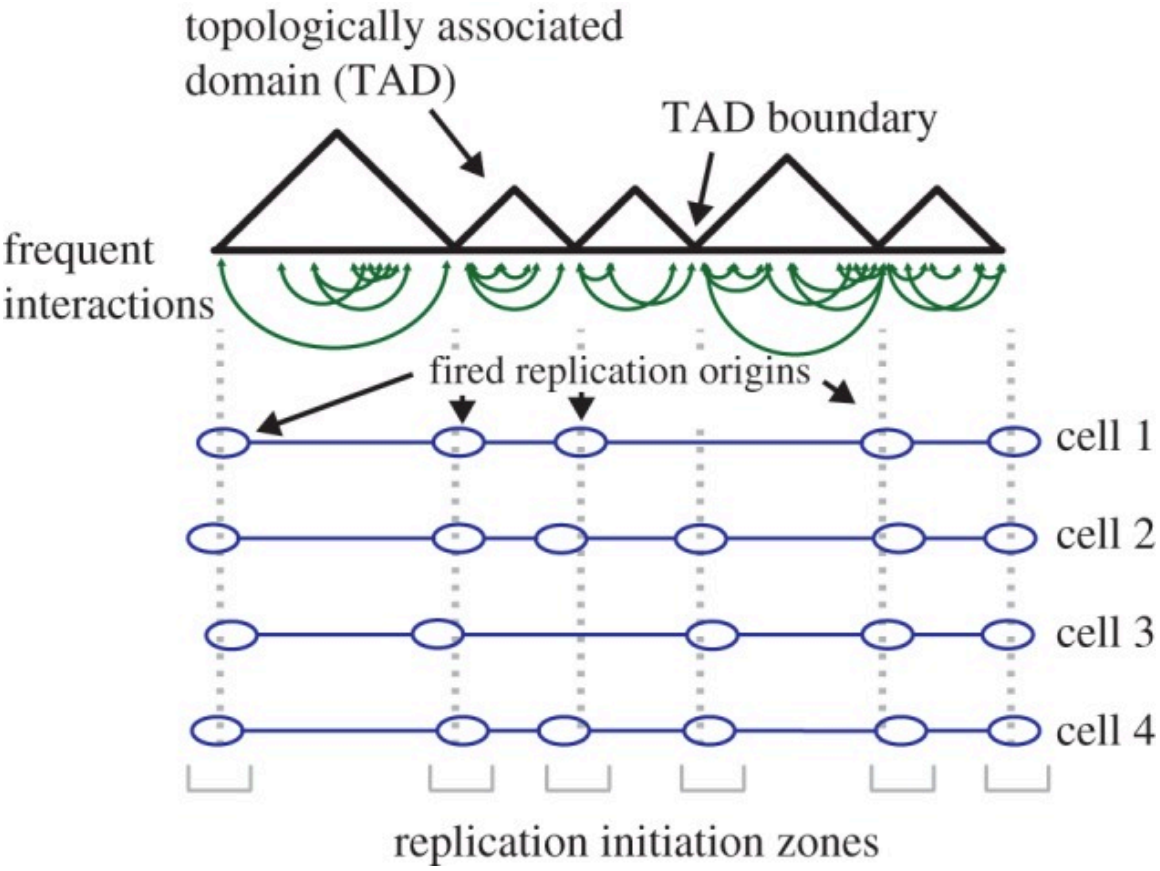

Figure 1.8. DNA replication origins are located in highly regulated boundaries between chromatin domains. Representation of a region of a chromosome containing topologically associating domains (TADs). Within TADs, chromatin elements make frequent associations with each other (depicted as green lines). TADs are separated by boundaries to limit interactions between different domains—enriched at TAD boundaries are replication origins (blue ovals) that initiate stochastically within these zones (depicted in grey).

that specific TADs displayed reproducible replication timing across cells based on their genomic compartment (Ryba et al. 2010, Pope et al. 2014). Empirical evidence suggests that the physical isolation of chromosomes into TADs functions in: 1) the coregulation of gene expression states and replication origins within the same TAD and 2) the exclusion of unspecified regulatory activity from adjacent TADs (Fanucchi et al. 2013, Phillips-Cremins et al. 2013, Dekker et al. 2016).

The segregation of these domains is thought to depend on stable boundary regions between adjacent TADs domains; metazoan TAD boundaries have been shown to be enriched for active chromatin marks, constitutively transcribed genes, DNA replication origins, and binding sites for chromatin insulator proteins such as CTCF (Ulianov et al. 2016, Dileep et al. 2015, Narendra et al. 2016). Two models have been proposed to explain the regulatory function of TAD boundaries, however the precise mechanism remains elusive (Nichols and Corces 2015). The first model proposes that the formation of loops which stabilize enhancer and promoter interactions is mediated by CTCF binding polarity; the binding of CTCF then recruits the ring-like SMC family protein complex Cohesin to help form cell-type specific loops between enhancers and target gene promoters (de Wit et al. 2015, Guo et al. 2015, Wendt et al. 2008, Kagey et al. 2010). Another model proposes that the additive effect of housekeeping gene transcription and CTCF binding at TAD boundaries generates a more rigid chromatin structure through the regular spaced arrays of nucleosomes, thus insulating the boundary and preventing interactions between

neighboring TADs (Vietri Rudan and Hadjur, 2015, Gaffney et al. 2012, Dixon et al. 2016).

Genome Organization: Operons

The spatial organization of genes along chromosomes influences the rate at which genes are transcribed and translated into proteins—the coregulation of multiple genes acting in the same pathway maximizes their expression efficiency and minimizes the regulatory code for transcription. This concept emerged after its was discovered that all the genes required for lactose utilization in *E. coli* were cotranscribed under the control of a single regulatory site—the transcription of these clusters (or polycistrons) from a single 5' promoter defines what are known as operons (Jacob et al. 1960). Despite their pervasive use in prokaryotes, operons are not common in eukaryotic genomes—instead, most eukaryotic genes are transcribed from individual promoters (Lawrence 1999). However, the fortuitous discovery of polycistronic gene clusters in the nematode *C. elegans* provided the first example of operon usage of the animal kingdom (Spieth et al. 1993).

Bacteria transcribe polycistronic mRNAs which are translated processively, as ribosomes re-initiate at each gene 5' end and terminate at the downstream 3' end (Leive and Kollin, 1967)—*C. elegans* transcribe polycistronic pre-mRNAs instead, which are terminated by 3' end formation and trans-spliced at each gene 5' end to make monocistronic mRNAs (Figure. 1.9A) (Blumenthal and Gleason 2003). Trans-splicing of mRNA 5' ends is common in nematodes, trypanosomes, cnidarian hydra,

and the primitive ascidian *Ciona intestinalis*—in the process, the 5' end of the RNA (spanning the promoter to the first 3' splice site) is replaced by a short 5' capped spliced leader (SL) (Bektesh and Hirsh, 1988; Sutton and Boothroyd,1986; Hastings, 2005). The trans-splicing reaction is a variant of cis-splicing, which co-transcriptionally removes introns from pre-mRNA transcripts (Blumenthal and Thomas, 1988). In *C. elegans* many of the same proteins catalyzing cis-splicing (U2, U4, U5, and U6 snRNPs) also are responsible for co-transcriptional trans-splicing (Zorio et al. 1994, Thomas et al. 1988). Analysis of *C. elegans* trans-spliced genes and their position within the polycistronic pre-mRNA revealed the existence of two types of SL snRNPs, the SL1 snRNP and the SL2 snRNP (Ferguson et al. 1996).

Spliced Leader Trans-splicing in *C. elegans*

Sequencing of 5' RNA ends suggest that SL1 snRNP is responsible for most trans-splicing events at the 5' end of polycistronic pre-mRNAs (Zorio et al. 1994, Allen et al. 2011) (Figure 1.9A). The finding that SL2 sequences were often found 100-300 bp downstream of another gene led to the proposal that SL2 trans-splicing is favored at downstream genes in operons (Huang and Hirsh, 1989, Allen et al. 2011. The trans-splicing efficiency of downstream mRNA transcripts relies to some degree on two DNA sequence elements: the AAUAAA 3' end cleavage signal, which is bound by the cleavage and polyadenylation specificity factor (CPSF) and a U-rich element in the linker region between cistrons, which is bound by cleavage stimulatory

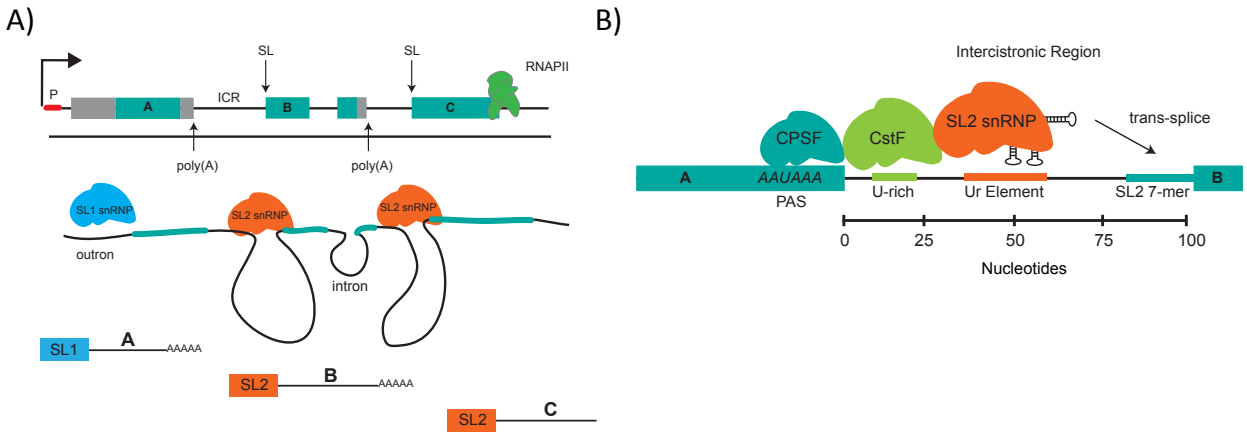

Figure 1.9. Operons and trans-splicing in *C. elegans.* A) Operonic genes (turquoise) are separated by an intercistronic region (ICR) approximately 100-300 bp in length– this ICR is flanked by a 3' poly A site in the upstream gene and a splice-leader (SL) acceptor site in the downstream gene. RNA polymerase (RNAP II) initiates synthesis from a single promoter (red), generating a polycistronic pre-mRNA that is bound by SL1 snRNPs (light blue) at the 5' promoter transcribed region (outron) and SL2 snRNPs (orange) at the 5' end downstream genes. Independent trans-splicing events are capable of producing three monocistronic mRNAs. B) Intercistronic recruitment of 3' end formation factors, CPSF and CsfF, at the polyA-site (PAS) (dark green) and the U-rich linker (light green), respectively, stabilizes the SL2 snRNP to U-containing sequences, thus minimizing the potential for 5-3' exonuclease digestion prior to protection by trans-splicing.

factor (CstF) (Figure 1.9B) (Kuersten et al. 1997, Evans et al. 2001, Liu et al. 2001). The presence of an intact cleavage signal on 3' end of an upstream gene increases the efficiency of SL2 trans-splicing, but is not absolutely required—however when the U-rich element is mutated, upstream 3' end formation can still occur but the downstream product is not produced, usually as a result of transcription termination (Huang et al. 2001, Kuersten et al. 1997). This data suggests that the two sequence elements and their associated factors act to coordinate 3' end formation by blocking potential 5'-3' exonuclease degradation of the downstream pre-mRNA until the SL2 snRNP is recruited to trans-splice the 5' end of the downstream transcript (Blumenthal 2013).

Most nascent transcripts emerging from operons are trans-spliced, however not all trans-spliced genes are contained in "classical" operons where genes are separated by ~100 bp (Boeck et al. 2016). Comparisons of operon structures and SL2 trans-spliced genes across nematode species suggest trans-splicing ability was present in a common ancestor and that operons developed early in nematode evolution due to a mechanism for downstream polycistronic pre-mRNA processing (Whitton et al. 2004).

Defining *C. elegans* Genes within Operons

In light of the discovery of splice leader sequences, microarray and RNA sequencing efforts sought to annotate worm genes contained within operons—current estimates are that 15-20% of *C. elegans* genes, corresponding to ~3,000-

4,000 genes, are contained within operons (Blumenthal 2013, Allen et al. 2011). Most operons contain an average of around 3 genes (longest is 8) and are found in gene-dense, recombination-low regions in the center of autosomes—operons are depleted on X-chromosomes (Blumenthal et al. 1998). A non-random arrangement of gene classes within operons also emerged from the analysis—genes encoding essential functions common to all eukaryotes were far more represented in operons than worm-specific proteins (Kamath et al. 2003). Genes contained in operons typically encoded proteins involved in RNA processing, gene expression, cellular metabolism, and protein homeostasis, while genes encoding cell type specific products such as homeodomain transcription factors, peptide transporters, and chemosensory modules were depleted from operons (Blumenthal et al. 1998).

Most organisms require potent control of gene expression regimes to respond to cell-intrinsic and extrinsic signals; however, only a few animal species other than nematodes organize their genes into operons (Lercher et al. 2003). The reason behind the exclusivity of operon structure to just a few animal species remains unknown—yet it appears likely that the capacity to process closely clustered genes through a mechanism such as trans-splicing would have preceded operon gene organization (Guiliano and Blaxter, 2006). Some have proposed that the co-regulation of operonic genes in *C. elegans* potentially reflects a design aimed at achieving more efficient, precise control over the expression of “viability” genes in response to a global signal, possibly starvation, extreme stress, or favorable environmental signals (Guiliano and Blaxter, 2006, Lercher et al. 2003). Given the

evolution and natural history of nematodes, such a suggestion would not be unreasonable—the arrangement of essential, viability genes into operons would provide potent control over the repression or silencing of growth genes during periods of starvation or food availability (Denver et al. 2005, Braendle et al. 2007).

Compartmentalization of the genome into TADs accomplishes a similar arrangement—clustering the necessary regulatory sites for transcription initiation concentrates limiting RNA Pol II cofactors on the expression of a subset of genes. The chromosome-conformation capture experiments that revealed extensive TADs during mammalian and Drosophila development revealed an absence of TAD organization on *C. elegans* chromosomes (Dixon et al. 2016). Holocentric chromosome structure and the lack of chromatin insulator sequences or proteins, like CTCF, might explain the lack of TAD structure in *C. elegans* genome organization (Rao et al. 2014). However, an integrated analysis of nascent transcription initiation and elongation patterns at specific developmental stages would be needed to provide critical information on how transcriptional activity is differentially regulated genome-wide.

While transcription is critical to pattern gene expression during embryogenesis, DNA replication often precedes zygotic transcriptional activation and is related to chromatin changes which occur during the early embryonic stages (Ke et al. 2017, Hug et al. 2017). The advent of sequencing approaches to profile DNA replication genome-wide has revealed that the regulation of DNA replication initiation is highly correlated with cell-type specific gene expression programs and three-

dimensional chromosome structure (Pope et al. 2014). The following section will introduce the critical regulatory concepts governing DNA replication during development.

Developmental Regulation of DNA Replication

In spite of the essential requirement to accurately copy genetic information, the core machinery that replicates the genome is widely diverged across all three domains of life: bacteria, archaea, and eukaryotes (O' Donnell et al. 2013). The tremendous level of plasticity, particularly observed among families of DNA polymerases, may reflect the unique nature of genome duplication—while more conserved molecular machines involved in translation and transcription can simply restart if the process fails, DNA replication happens only once in the lifetime of a cell and must succeed for the propagation of the genome (Yao and O'Donnell, 2016). DNA polymerases and their accessory factors differ in their capacity to traverse numerous DNA sequences, lesions, or DNA binding proteins—thus, genomes have adapted by evolving replication strategies which use exchangeable polymerases that can simultaneously copy both DNA strands with high fidelity.

James Watson and Francis Crick observed in their landmark model of DNA that the antiparallel double helix must be unwound and separated in order to template the formation of a new companion chain that is complementary in sequence (Watson and Crick, 1953). However, their structure revealed that if semiconservative replication is to take place simultaneously, as it does in eukaryotes, the strands must

be synthesized in opposite directions—yet, the nucleotide precursors that initiate synthesis are only activated on the 5' end (Yao and O'Donnell, 2016). So how did the complex architecture of DNA replication machines evolve to efficiently replicate the eukaryotic genome? Structural studies have indicated that enzymes such as helicases, primases, and sliding clamps must have evolved within specific genomic and developmental contexts during eukaryotic evolution to ensure timely replication of the genome sequence (Yao and O'Donnell, 2016). Another point of consideration in eukaryotes is the packaging of DNA strands into chromatin, which mandated that replication machines must be integrated with enzymes that remodel, modify, and assemble nucleosomes during S-phase (O' Donnell et al. 2013).

While many aspects of genome duplication are active areas of inquiry, studies of DNA replication across eukaryotes has benefitted immensely from a variety of creative assays to measure DNA replication dynamics, ranging from reconstituted *in vitro* replication with purified proteins to high-resolution genomic maps of DNA replication from next-generation sequencing data. This section will provide a summary of key regulatory steps governing the initiation of DNA replication across the genome before transitioning to a review of how chromatin and gene activity influence the timing with which DNA replication origins are used during S-phase.

Initiation at the Origin

Replication initiation is regulated at origins, which are discrete genomic sites bound by initiator proteins prior to DNA synthesis. The number of origins and the

speed of the replication fork determine the time it takes to replicate the genome (Stillman 2005). In the bacterium *E. coli*, replicating from a single origin at a rate of 1 kilobase of DNA per second allows for duplication of the genome in 30 minutes (Kim et al. 1996). In eukaryotes, fork rate is 20 times slower and requires the use of many replication origins along the linear chromosome (Rivin and Fangman, 1980).

Only a subset of possible replication origins will fire during S-phase—the likelihood that an origin will fire in a cell, or its efficiency, has been directly correlated with its chromatin context and the developmental state of the cell (Gilbert et al. 2010). The activation of DNA replication origins requires two discrete steps: origin licensing in G1 phase and origin firing in S-phase, both regulated by the AAA+ family of multimeric initiator proteins and translocases (Erzberger and Berger 2006). The eukaryotic initiator, referred to as the origin recognition complex (ORC), is a hexameric, ring-shaped protein composed of five ORC subunits and another AAA+ protein called Cdc6—ORC is responsible for loading the two replicative helicases that will give rise to two diverging replication forks once the origin is fired (Bell and Stillman, 1992, Cocker et al. 1996). The replicative helicase is generally conserved in form and function across all domains of life—represented by a six subunit, ATP-dependent DNA translocase, the primary responsibility of the helicase is to physically separate the DNA duplex to gain access to each template strand (Schwacha and Bell, 2001).

A series of accessory proteins and regulatory transitions ultimately transform the inactive MCM2-7 hexamers, called the pre-replicative complex (pre-RC) into the

active helicase complex, called the CMG (Cdc45-MCM2-7-GINS) (Remus et al. 2009, Georgescu et al. 2017). The final steps of activation, through assembly of Cdc-45 and GINS into the pre-RC, are regulated by two cell cycle regulated kinases, Dbf4-dependent kinase (DDK) and cyclin-dependent kinase (CDK) (Tanaka et al. 2007, Heller et al. 2011).

The core components of the replisome include DNA polymerases, circular sliding clamps, clamp loaders, helicase, primase, and single stranded binding protein (SSB) (Waga and Stillman 1998, Johnson and O'Donnell 2005, Barry and Bell 2006). The anatomy of the replication fork is asymmetric—obligate DNA synthesis in the 5' to 3' direction requires that polymerases and processivity factors synthesize the leading strand continuously in the direction of fork movement and the lagging strand is synthesized discontinuously as Okazaki fragments (OFs), in opposite direction of fork movement (Anderson and DePamphilis, 1979). Eukaryotic replication relies on three polymerases—Pol ε as the leading strand polymerase, Pol δ as the lagging strand polymerase, and Pol α/primase which synthesizes the start of each OF on the lagging strand (Tsurimoto and Stillman, 1991, McElhinny et al. 2008).

While much of this work to describe the enzymology of the replisome at each activation step was performed *in vitro* on a naked DNA template, the organization of genomes into chromatin profoundly affects DNA replication dynamics (Li and Kelly, 1984, Groth et al. 2007, Tan et al. 2006). The following section details how DNA replication proceeds through the chromatin landscape in the nucleus and highlights

the role for DNA replication in preserving chromatin states at promoters, thus, propagating gene expression states through cell divisions.

DNA Replication through a Chromatin Landscape

In spite of an abundance of origins spread throughout the genome, normal cells often fire replication origins in a surprisingly consistent timing program, suggesting replication is integrated into cell regulatory states (Davis et al. 2001, Dimitrova and Gilbert, 1999). Studying the highly reproducible origin firing sequence along budding yeast chromosomes revealed two timing groups: early and late firing origins (Raghuraman et al. 2001). In spite of an apparent fixed timing of early firing and late firing origins in budding yeast, simply moving a late replicating origin to an early replicating region of the chromosome advances its time of firing (Vogelauer et al. 2002). Altering the chromatin context by specifically targeting histone acetyltransferases or deacetylases can similarly enhance or suppress the probability of initiation for local replication origins (Vogelauer et al. 2002, Goren et al. 2008).

Genome-wide surveys of ORC binding across different cell types reveals that origin selection in metazoans is influenced by the developmental state of the cell, as ORC typically overlaps with sites of RNA polymerase II occupancy, transcription factor binding sites, and CpG island promoters (Macalpine et al. 2004, Cadoret et al. 2008). Studies of replication initiation in Xenopus egg extracts revealed random patterns of initiation on a DNA template until the addition of a transcription domain led to site-specific initiation and hyperacetylation of the origin (Danis et al. 2004),

suggesting that the assembly of transcription complexes at regulatory sites may directly or indirectly facilitate the loading of pre-RCs (Danis et al. 2004).

Studies in the Gilbert lab demonstrated how vertebrate cells organize their chromosomes into broad replication domains to integrate the timing of replication into the regulatory state of the cell (Rivera-Mulia and Gilbert 2016). Replication domains in vertebrates can be divided into two classes: constant timing regions (CTRs) comprising domains of clustered origins with consistent firing patterns from cell to cell, and timing transition regions (TTRs) representing regions which are passively replicated later in S-phase (Pope et al. 2014). Surprisingly, the CTRs tend to overlap with constitutive TAD boundaries across various cell types—these boundaries harbor "housekeeping genes" consistently expressed in cycling cells (Pope et al. 2014, Rudan et al. 2015).

Replication and the Stable Inheritance of Chromatin States

The timely completion of DNA replication and the inheritance of chromatin states are essential for cell viability—therefore disruption of the parental chromatin structure and reassembly of the cognate structure on newly replicated daughter genomes requires extensive regulation (Moazed 2011). *In vitro* replication systems have identified several chromatin factors, including histone chaperones (FACT, Nhp6, Asf1), chromatin remodeling complexes (INO80, Isw1), and histone acetyltransferase complexes (NuA4, SAGA), that facilitate the passage of the replisome through chromatin (Devbhandari et al. 2017, Kurat et al. 2017).

A longstanding question, however, has focused on the basic mechanisms that propagate chromatin states after passage of the replication fork (Alabert and Groth 2012). Studies from Drosophila cells and budding yeast showed increased nucleosome occupancy around active promoter regions and enhancer regions shortly after DNA replication, suggesting a possible competition between nucleosomes and transcription factors in the wake of the replication fork (Fennessy and Owen-Hughes 2016, Ramachandran and Henikoff 2016). A study from the Groth lab that quantitatively tracked parentally modified histones and nascent histones deposited in the wake of the replication fork, revealed that histone modification patterns (in repressed and active genomic regions) are preserved during replication, with parental modified histones being deposited within 250 bp of their pre-replication position (Reveron-Gomez et al. 2018). Interestingly, the genome-wide kinetics with which PTM patterns were restored depended on the local enrichment of pre-modified histones and their associated enzymes. PTMs associated with transcribed promoters, such as H3K4me2, were rapidly restored by G2 phase of the same cell cycle—H3K27me3 domains were restored by the cell cycle following DNA replication, however the domains with the higher EZH2 methyltransferase occupancy were restored by 10 hours post replication, compared to 24 hours in low EZH2 domains (Reveron-Gomez et al. 2018, Alabert et al. 2015, Xu et al. 2012).

In light of this evidence, the engagement of common chromatin factors by DNA replication and transcription machinery appears to coordinate the stable propagation of the gene regulatory state through cell divisions (Yadav and

Whitehouse, 2016, Coleman and Struhl, 2017) (Figure 1.2.1). This model is supported by evidence that DNA replication in *Drosophila* embryos provides an opportunity for the transcription factors Zelda and GAGA-factor to bind their enhancers and establish stable chromatin accessibility patterns prior to zygotic genome activation (Blythe and Wieschaus, 2016).

Main Goals of this Thesis

Decades of classical genetic studies in *C. elegans* have revealed the conserved regulatory pathways that pattern the embryo. However a mechanistic understanding of how the organization of the genome influences the principle mechanisms of DNA replication, gene transcription, and chromatin organization remain largely unknown. The modest goal of this thesis is to apply basic genomics approaches in order to record the dynamics of chromatin, DNA replication, and transcription landscapes in the developing worm embryo. The connections between these genome transactions places a premium importance on understanding how these distinct processes are established during early embryogenesis. Our approach was to generate simple maps of the genomic landscape at critical stages of *C. elegans* embryogenesis—we hoped that by overlaying maps of gene transcription and DNA replication at identical embryonic stages, we might be able to distinguish unifying patterns and organizing principles of the *C. elegans* genome which patterning the entire cell lineage of the animal.

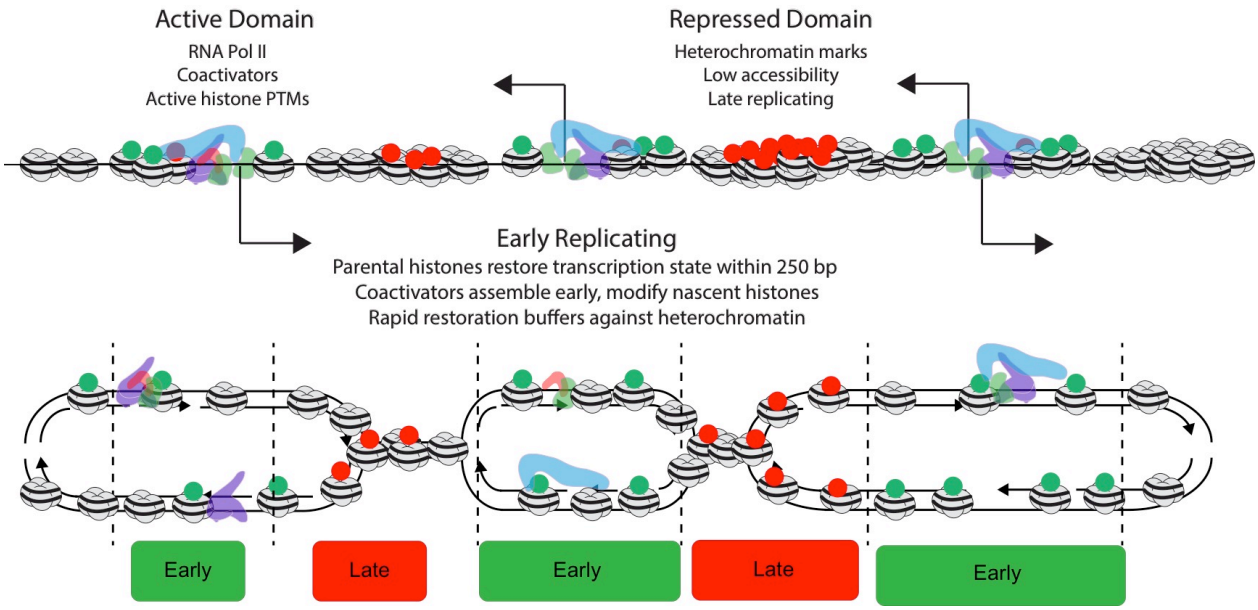

Figure 1.2.1. Chromatin assembly at DNA replication origins propagates transcriptional memory during cell division. Actively transcribed domains (arrows indicating transcription), comprising RNA Pol II (green), coactivators such as SAGA/CBP (purple) and Mediator (blue), and active histone marks, such as H3K27ac (green dots), are replicated early in S-phase based on their proximity to efficient DNA replication origins. Repressed domains are passively replicated by distant origins, resulting in their replication late in S-phase, and are characterized by histone marks associated with compact heterochromatin (red dots). The rapid reassembly of active domain proteins and nucleosomes, temporally buffers against repressive factors spreading to transcribed promoters, thus preserving gene expression states.

While DNA replication profiles have been described for several metazoan species, pinpointing the factors which specify replication origin activity in various animal cell types remains an active area of investigation. Considering the wealth of information on *C. elegans* embryo development and the unique regulation of cell lineage specification in the early embryo, the first aim of this work was to address how the *C. elegans* DNA replication landscape becomes specified and coordinated in embryonic time and space.

The regulation of the DNA replication landscape has been shown to be influenced by several determinants of eukaryotic chromosome structure, including chromatin accessibility and gene transcription. In the course of *C. elegans* embryogenesis, diverse gene expression programs must be coordinately activated to ensure proper patterning of cell fate and tissue function—the second aim of this work was to identify how gene transcription is regulated during the embryonic developmental transitions.

The final aim of this thesis is to understand how these two processes of DNA replication and gene transcription are harmonized during *C. elegans* embryogenesis. The activation of the zygotic genome has been shown to profoundly influence DNA replication and cell cycle dynamics during *Drosophila*, *Xenopus,* and zebrafish embryogenesis, supposedly through a limitation in the number of DNA replication origins which are fired. Our fundamental questions, which we applied to a model of *C. elegans* embryogenesis, relate to:

- How many DNA replication origins are used throughout embryonic development?
- Do DNA replication origins require zygotic transcription for specification?
- Is origin activity constant throughout embryo development or does it change with a shifting gene expression landscape?
- What happens to DNA replication origin domains in terminally differentiated cells?
- How does chromosome structure or modification state change in response to dynamic gene expression programs over the course of embryonic development?

The work described herein will hopefully contribute useful datasets and analysis that can advance our understanding of *C. elegans* embryogenesis the evolution of metazoan genome structures.

Chapter 2

DNA Replication in the *C. elegans* Embryo

Introduction

Origin Specification in Metazoan Embryos

One of the initial tasks for developing embryos is genome replication, a process that requires the regulated licensing and initiation of many origins of replication along the chromosomes (O'Donnell et al. 2013). As was discussed in Chapter 1, numerous factors determine where DNA replication origins are specified and the time at which they are active during S-phase. Studies in metazoa have found particular chromatin signatures that typify transcriptionally active chromatin are enriched at replication initiation sites, suggesting that accessible chromatin regions, established near transcribed genes, may also guide the assembly of pre-RC complexes (Danis et al. 2004). The number of DNA replication origins which are activated within embryonic cells also appears to be developmentally controlled (Blumenthal et al. 1974; Foe & Alberts, 1983). The rapid genome duplication observed during early cleavage cycles in *Drosophila* and *Xenopus* embryos is facilitated by the use of many, closely spaced, replication initiation sites; however, the onset of zygotic transcription is accompanied by a reduction in the number of replication initiation events, causing a marked increase in S-phase length (Blumenthal et al. 1974; Foe & Alberts, 1983).

To address the fundamental question of how DNA replication is integrated with the profound cellular changes that occur during embryo development, we chose to investigate DNA replication within developing *C. elegans* embryos. Our goal was to characterize the signatures of replication origins and develop an understanding of replication patterns through early development. The work presented in Chapter 2 was performed in collaboration with Dr. Ehsan Pourkarimi and the results of those experiments have been published (Pourkarimi et al. 2016). Dr. Pourkarimi is credited with the original concept, experimentation, data analysis, and primary authorship. James Bellush is credited on the publication with experimentation, data analysis, and contributing authorship. Data and figures which have been recapitulated from Pourkarimi et al. 2016 will be denoted as such in the text.

DNA Replication and *C. elegans* Lineage Establishment

The invariant *C. elegans* cell lineage depends upon a reproducible pattern of founder cell divisions that initiate in the early embryo (Scierrenberg, 2016). The early cell division cycles guide the developmental patterning of the animal through two principal mechanisms; the asymmetric segregation of maternal RNAs and the consistent cell-cell interactions that spatially organize the embryonic body plan (Strome, 1986; Hutter & Schnabel, 1994; Mello et al. 1992). Analysis of these invariant cell division patterns has provided valuable evidence of the coordination between cell proliferation and fate differentiation in the *C. elegans* embryo (Bao et al. 2008).

By analyzing the appearance of cell fate markers relative to daughter cell divisions, Bao et al. found that in the early embryo, the cell cycle pace is established by the founder cell fate and is subsequently fine-tuned by tissue and organ differentiation within each lineage (Bao et al. 2008). It remains unclear how cell fate is specified during the rapid embryonic divisions, however the concerted regulation of DNA replication and cell cycle factors has been shown to be essential for embryo viability (Hebeisen & Roy, 2008; Edgar & Wood, 1988). In a seminal study examining the relationship between DNA replication and lineage-specification in the early *C. elegans* embryo, the Wood group used aphidicolin to acutely inhibit DNA synthesis before or after the clonal establishment of the endoderm lineage in 8-cell embryos (Edgar & Wood, 1988). By monitoring the expression of two endoderm-specific markers, the authors found that inhibition of DNA synthesis prior to the first cell cycle after lineage establishment prevented marker expression; however, embryos which were treated with aphidicolin after permitting a short period of DNA synthesis showed robust marker expression (Edgar & Wood, 1988). Altogether, these studies suggest that the expression of tissue-specific factors in early *C. elegans* embryos is not dependent upon cell cycle number and S-phase length, as has been shown for *Drosophila* and *Xenopus* embryos; instead, lineage-specific gene expression programs seem to be coupled to active DNA synthesis within clonally established cells (Edgar & Wood, 1988; Farrell & O'Farrell, 2013; Kane & Kimmel, 1993; Aladjem, 2007).

Chromatin Signatures of Metazoan Replication Origins

The metastability of histone modifications imparts a level of dynamic control to the embryonic transcriptome; chromatin signatures corresponding to particular transcriptional states can be propagated across multiple cell divisions or they can be erased and remodeled according to the appropriate developmental state of the cell (Jenuwein & Allis, 2001, Bernstein et al. 2001, Alabert & Groth, 2012). Interestingly, many of the combinatorial patterns of histone marks associated with gene transcription and DNA replication origins are conserved across metazoa, suggesting a shared mechanism of specification by chromatin (Ho et al. 2014; Lubelsky et al. 2014). Histone acetylation and open chromatin structures have been associated with origin specification by ORC in both *Drosophila* and human cells (Lubelsky et al. 2014, Iizuka & Stillman, 1999). Additional studies suggest that hyperacetylation of nucleosomes at replication origins is important to regulate origin activity during endoreplication in *Drosophila* follicle cells and during gastrulation in *Xenopus* embryos*;* furthermore, hyperacetylation in both contexts correlates to an increase both in gene transcription and DNA replication origin activity (Almouzni et al. 1994; Aggarwal & Calvi, 2004).

Mapping DNA Replication Origins with Okazaki Fragment Sequencing

Our stated goal of recording DNA replication dynamics during early embryo development relies on a method to precisely identify origin locations and activity. Studies of DNA replication in the budding yeast *S. cerevisiae* have pioneered novel

sequencing techniques to analyze DNA replication patterns, however one inherent challenge is deducing replication signatures from a population of cells growing at steady state (Raghuraman et al. 2001, Wyrick et al. 2001, Nieduszynski et al. 2006, Smith and Whitehouse, 2012).

The development of a method to purify and deep sequence Okazaki fragments (OFs), the lagging strand products of replication forks, have provided high-resolution maps of DNA replication across budding yeast and human genomes (Smith & Whitehouse, 2012; McGuffee et al. 2013; Petryk et al. 2016). OFs are synthesized discontinuously on the lagging strand DNA polymerase δ and mature through a unique process of strand displacement synthesis and 5' end trimming by the structure specific endonucleases Fen1 and Dna2; after cleavage of OF 5' ends, they are ligated into a continuous strand by DNA ligase I (Anderson and Depamphilis, 1979, Bae and Seo, 2000, Levin et al. 1997). Because OFs provide strand and positional information of active DNA replication forks, a strategy to map replication origins was designed on the assumption that transient depletion of DNA ligase I, through an inducible degron, would enrich for OFs in S-phase cells—the single stranded fragments are purified, ligated, and sequenced to provide a "snap-shot" of replication forks that can be measured by their strand specific coverage across chromosomes (Smith and Whitehouse 2012).

The key to the strategy is the preservation of strand identity, which allows OF density to be measured at replication forks throughout the genome (Figure 2.1A) (McGuffee et al. 2013). The origins positions are readily detected as sharp

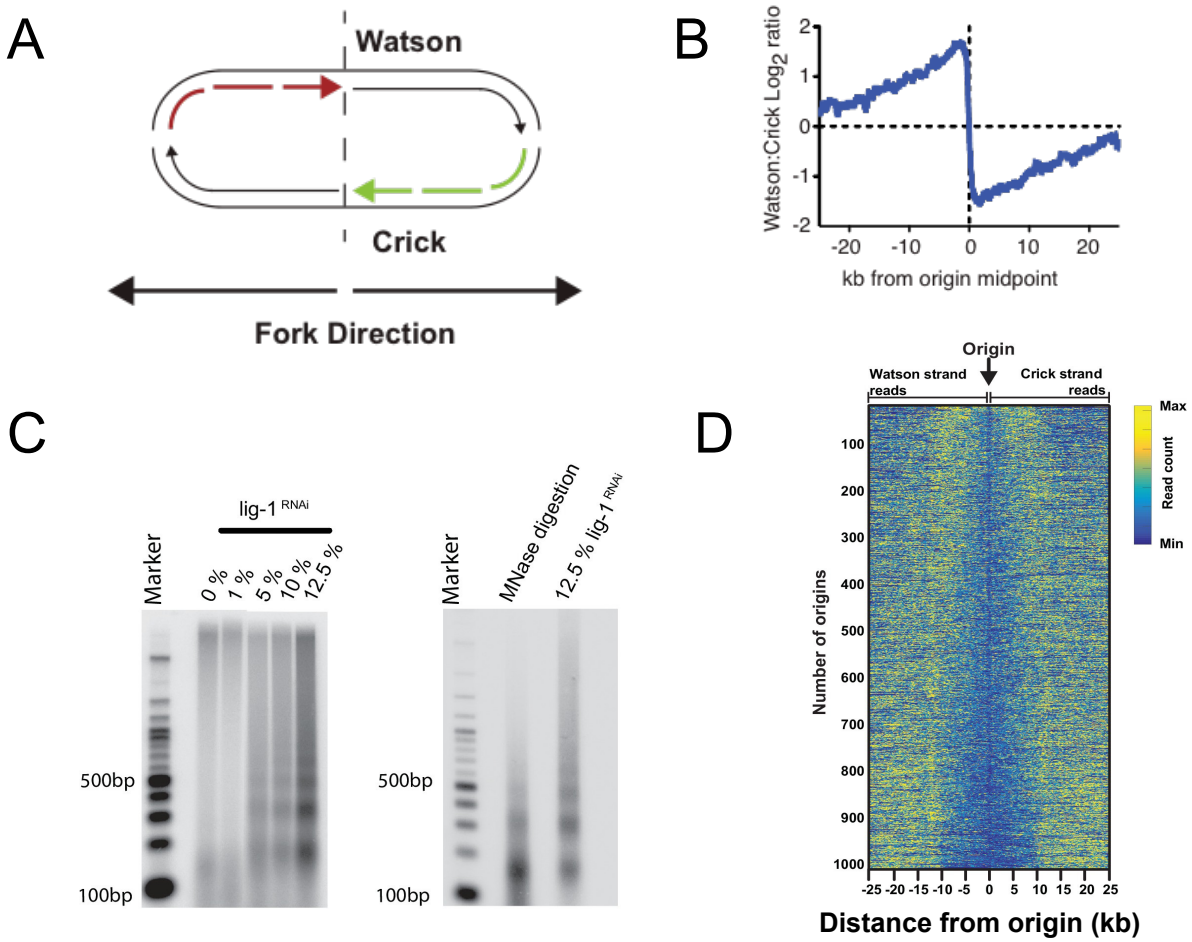

Figure 2.1: Mapping DNA replication origins in *C. elegans* embryos. A) The anticipated distribution of Okazaki fragments surrounding an efficient DNA origin. B) Strand specificity of Okazaki fragment sequencing was validated by mapping the Log_2 ratio of Watson:Crick fragment reads around confirmed replication origins identified in budding yeast (Smith and Whitehouse, 2012). C) Left panel: increasing DNA ligase I (*lig-1*) depletion by RNAi results in the generation of small DNA molecules with periodic size distribution. Right panel: comparison of Okazaki fragments with DNA resultant from a Micrococcal Nuclease digestion of chromatin, revealing similar periodicity. D) Variation among DNA replication initiation zones in *C. elegans* embryos. Data from Watson (left-moving fork) or Crick (right-moving fork) reads at efficient replication origins (n=1000, efficiency>0.5). Watson or Crick reads were mapped ±25 kb from each origin centered on the heatmap; the reads were normalized such that the sum of the squares of the values = 1. Origins ranked by the size of their transition zone, smallest at top.

transitions in leftward and rightward-moving replication forks—the magnitude of the transition at the origin is directly proportional to its efficiency, or likelihood the origin is used during S-phase (McGuffee et al. 2013) (Figure 2.1B). The rigor of this quantitative approach was independently confirmed by testing origin measurements against previously reported replication origin locations and firing times—measuring the density of OFs in four-part sliding windows across the genome was able to reproducibly determine known origin positions and firing times from ChIP of ORC or MCM, 2-D gels, or pulse-labeling approaches (Xue et al. 2006, Yashamita et al. 1997, Nieduszynski et al. 2007).

Results

Okazaki Fragment Sequencing in *C. elegans* Embryos

Harvesting a sufficient quantity OFs from *C. elegans* embryos required transient depletion of the DNA ligase involved in OF ligation, *lig-1* (Figure 2.1C). To minimize genome instability and prevent embryonic lethality associated with complete *lig-1* depletion, we determined an optimal level of depletion to maintain normal development, while enriching for fragments by dilution of the *lig-1* RNAi bacteria. Resolution of labeled DNA harvested from *lig-1* depleted embryos on a denaturing agarose gel, revealed a highly periodic pattern of OF size, similar to that of the nucleosome repeat uncovered by MNase digestion (Figure 2.1C) (McGuffee et al. 2013, Yadav et al. 2016). This pattern arises due the replication-coupled assembly of nucleosomes on nascent chromatin, causing polymerase δ and Fen1

endonucleases to generate OFs of heterogeneous lengths corresponding to nucleosome spacing (Smith and Whitehouse, 2012).

The fragments that mapped to the WS220 worm reference genome displayed a clear strand bias, which matches the expected distribution of OFs synthesized on the lagging strand: replication origins are at sites of transition from Watson to Crick strand reads (Figure 2.1D; Figure 2.2A). While the data is noisier than budding yeast – in part due to the high A:T content of the genome, residual DNA ligase activity and the complexity of the sample preparation – we identified >2000 replication origins using a modified mapping protocol. Median spacing between origins is 40 kb (Figure 2.2B), and 96% of origins are within 100 kb of another origin; the median efficiency – that is, the likelihood that a given origin is utilized during S phase, is ~50%, although we note that numerous origins are fired in most cells in the population. In contrast to the discrete initiation patterns around origins in budding yeast, replication in *C. elegans* apparently initiates within broad zones at most origins, similar to human cells (Figure 2.1D; Figure 2.2A) (Petryk et al. 2016, Dijkwel et al. 1991). Replication origins potentially influence gene organization: genes near origins tend to be shorter (Figure 2.2C) and genes flanking origins show a pronounced bias to ensure that replication fork progression and transcription are co-directional (Figure 2.2C-D).

Replication Origins Defined by Chromatin Marks of Transcription

Having generated a high-resolution map of replication origins, we next considered whether particular chromatin signatures are enriched near origins. Thus,

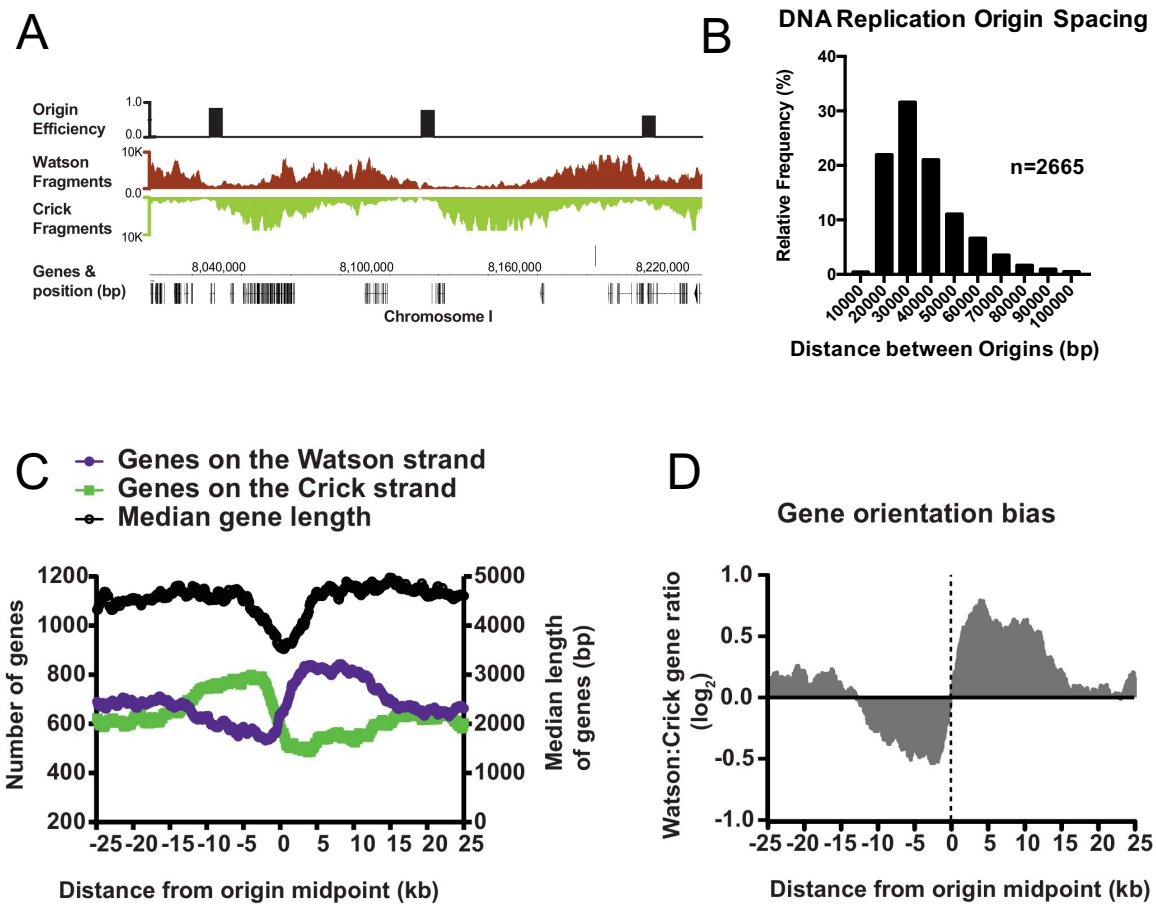

Figure 2.2: Genomic features of DNA replication origins. A) Fragments mapped to the *C. elegans* WS220 genome display a characteristic strand bias expected of Okazaki fragments. Regions of high Watson-strand signal (red) that transition to regions of Crick-strand signal (green) are replication origins (black bars). B) Frequency distribution (%) of DNA replication origin spacing across the *C. elegans* genome. Inter-origin distances were calculated using the midpoints of 2665 DNA replication origins from early stage *C. elegans* embryos; origin distances were binned into ten genomic windows using Prism and the relative percentage of origin distances in each bin are represented on the Y-axis. The center of each window is represented on the X-axis. C) Gene length and strand bias of genes within ±25 kb of replication origins. Median gene length (translation start to end) is shown by the black line on the right axis. Strand bias of the gene orientation for Crick strand genes (green) and Watson strand genes (purple), with respect to the template strand. D) Gene orientation bias from B is displayed as a log2 ratio of Watson : Crick genes.

we analyzed the abundance of several histone PTMs that were mapped in embryos at similar stages (Gerstein et al. 2010). Most of the embryos profiled for histone PTMs in the ModEncode datasets were mixed early stage embryos, spanning approximately 10-60 cells (Evans et al. 2016, Ho et al. 2014). Comparison of published histone PTM profiles with origin positions revealed remarkable concordance between the location of replication origins and acetylation of histone H3 and methylation of H3 at lysine 4 (K4) (Figure 2.3A). Analyzing H3K27ac (H3K4me2 is near identical), we find that 75% of efficient origins (>30% efficiency) are within 1 kb of the histone mark and most H3K27ac peaks overlap origins (Figure 2.3B). By surveying the abundance of various modifications present at replication origins, we tested whether other modifications behaved similarly to H3K27ac— this revealed that acetylation of H3 at K18, K23, K27 and H3K4me2 are not only present at origins, but the level of modification increases according to origin efficiency (Figure 2.3C). Significantly the histone modifications we find most strongly associated with origins are those that define gene enhancers in metazoan genomes (Ho et al. 2014).

Origin Firing Pattern Independent of Zygotic Transcription

Given that mixed stage embryos (28-150 cells) are transcriptionally active, our finding that replication initiates near histone marks linked with enhancers indicates that origins are localized to ‘accessible’ transcriptionally active regions. In this scenario, replication origins become defined once zygotic transcription has initiated; DNA replication in pre-gastrula embryos (PG), where transcription is limited (Edgar et

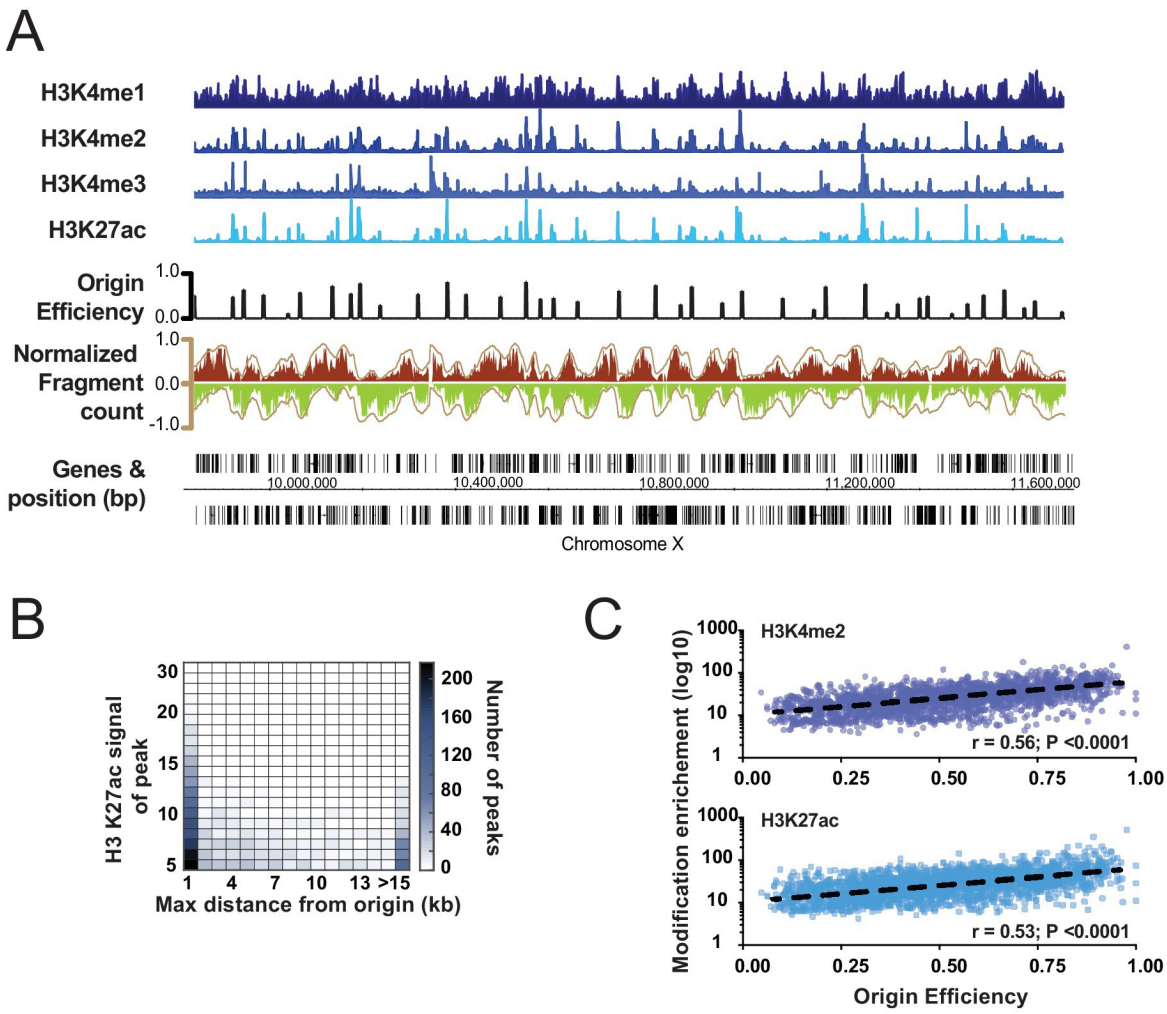

Figure 2.3: Correlation of select histone marks with origins. A) ChIP signal for active histone marks shown for a 1 Mb region of Chromsome X. Mapped origins (black) are coincident with peaks in the ChIP signal. B) Correlation between H3K27ac and replication origins is displayed as a heat map; origins and peaks were defined and measured computationally (Methods). Plotted are the number of peaks (left axis) within the maximum distance from the origin (X-axis); the relative intensity represents the fold enrichment over background and is plotted on the Y-axis. C) ChIP signal scales with increasing origin efficiency; level of histone modifications at replication origins is plotted for H3K27ac and H3K4me2 relative to the efficiency of the origin.

al. 1994), would presumably initiate from many, seemingly random sites—much like the apparent situation in *Drosophila* and *Xenopus* embryos prior to MBT (Robertson and Lin 2015). To investigate this, we harvested pre-gastrula embryos from a hyperactive egg laying strain (PG: median 9, mean 13 cells/embryo, n = 50; and late embryos from wild-type worms (L: ~200–558 cells/embryo), which represent different levels and extent of gene expression (Figure 2.4A-B). We compared origin location and efficiency for each population and found that the patterns of replication were highly similar, especially for the most efficient origins (Figure 2.4C-D). Evidently, even though gene transcription is fundamentally altered through this time course, replication origin usage is globally similar through early development (Figure 2.4D).

The establishment of specific replication origins in *C. elegans* occurs prior to the broad onset of zygotic transcription (Figure 2.4C; PG embryos)—this result supports findings by Wood and colleagues that embryos can divide up to the 100 cell stage in the absence of zygotic RNA polymerase activity (Edgar et al. 1994). Indeed, S phase length, lineage-specific cell cycle timing, asymmetric cell division and cleavage patterns all proceed normally in the absence of transcription in pre-gastrula early embryos—suggesting that the DNA replication program and early embryonic divisions are buffered against dramatic changes in zygotic transcription (Robertson and Lin 2015, Edgar and McGhee 1988).

Dynamic Remodeling of Transcription with Respect to Replication Origins

Coupling of DNA replication with gene enhancers within rapidly dividing embryonic cells likely imposes particular constraints upon genome duplication. S-

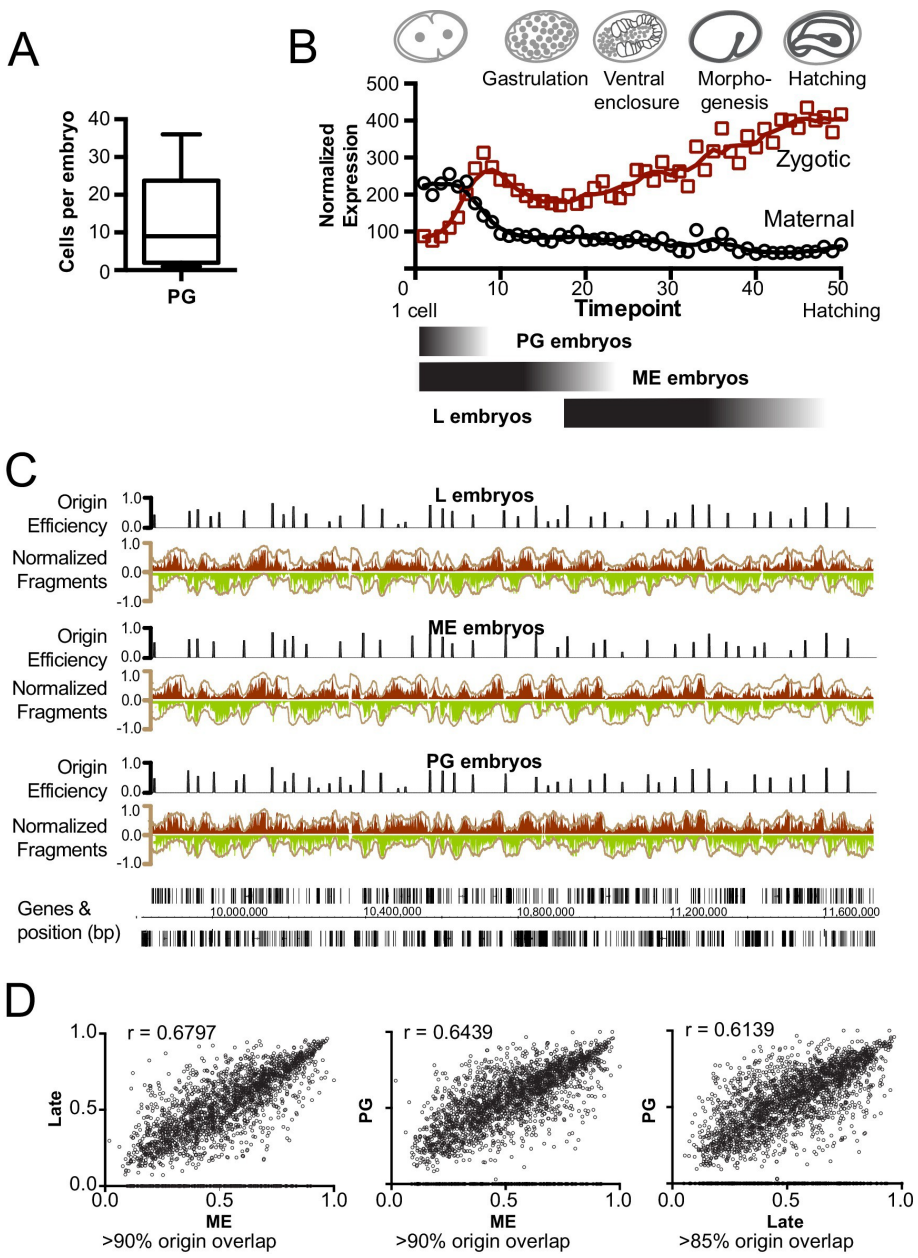

Figure 2.4: Mapping replication origins through early embryogenesis. A) Box plot showing the number of cells/embryo for a representative sample of pre-gastrula (PG) embryos, whiskers indicate the range of values. B) Graph showing maternal and zygotic transcripts through embryogenesis; data from Hashimony et al. 2015 were normalized such that the sum of expression for each gene = 1; total transcript levels for all genes at the timepoint are shown. General stages of embryos represented by black bars below: late (L), mixed early (ME), and pre-gastrula (PG). C) Okazaki fragment sequencing reveals broadly similar replication pattern. D) Scatterplot comparing origin efficiencies of overlapping origins within the L, ME, and PG datasets. Spearman correlation (r) is shown on graph. The percent origins overlap spatially is shown below.

phase in pre-gastrula embryonic cells ranges from ~10 to 25 min, depending on the lineage (Edgar and McGhee 1988). In an ideal system, where all origins fire in each S phase (100% efficiency), complete genome duplication could be achieved in ~15 min if origins are uniformly spaced every ~75 kb and replication forks progress at 2.5 kb/min (Blumenthal et al. 1974). We find that the median origin spacing is ~40 kb, which reflects that not all origins are fired within each S phase. Given that DNA replication appears to initiate at enhancers (Figure 2.3A-B), the requirement for closely spaced origins to achieve timely genome duplication would seemingly constrain the positions of enhancers across the genome. Thus, enhancers, and the genes they regulate, may be organized to facilitate the execution of the DNA replication and transcription programs in rapidly dividing embryonic cells.

To investigate how DNA replication origins in *C. elegans* embryos are positioned relative to gene transcription, we compared origin positions with the whole embryo transcriptome time series from individual embryos, spanning the one-cell stage through hatching (Hashimony et al. 2015). For each of the 50 time points, we removed maternally derived transcripts and plotted the sum of normalized transcript levels for each gene relative to the midpoint of efficient replication origins (Figure 2.5). We observe that mRNA transcript abundance steadily increases, until time point ~9 and coincident with gastrulation, there is an apparent clustering of actively transcribed genes near replication origins. The association of transcription with origins spans until time point ~25 and broadly scales with origin efficiency. However, beyond time point 25, the association breaks down—as gene expression changes,

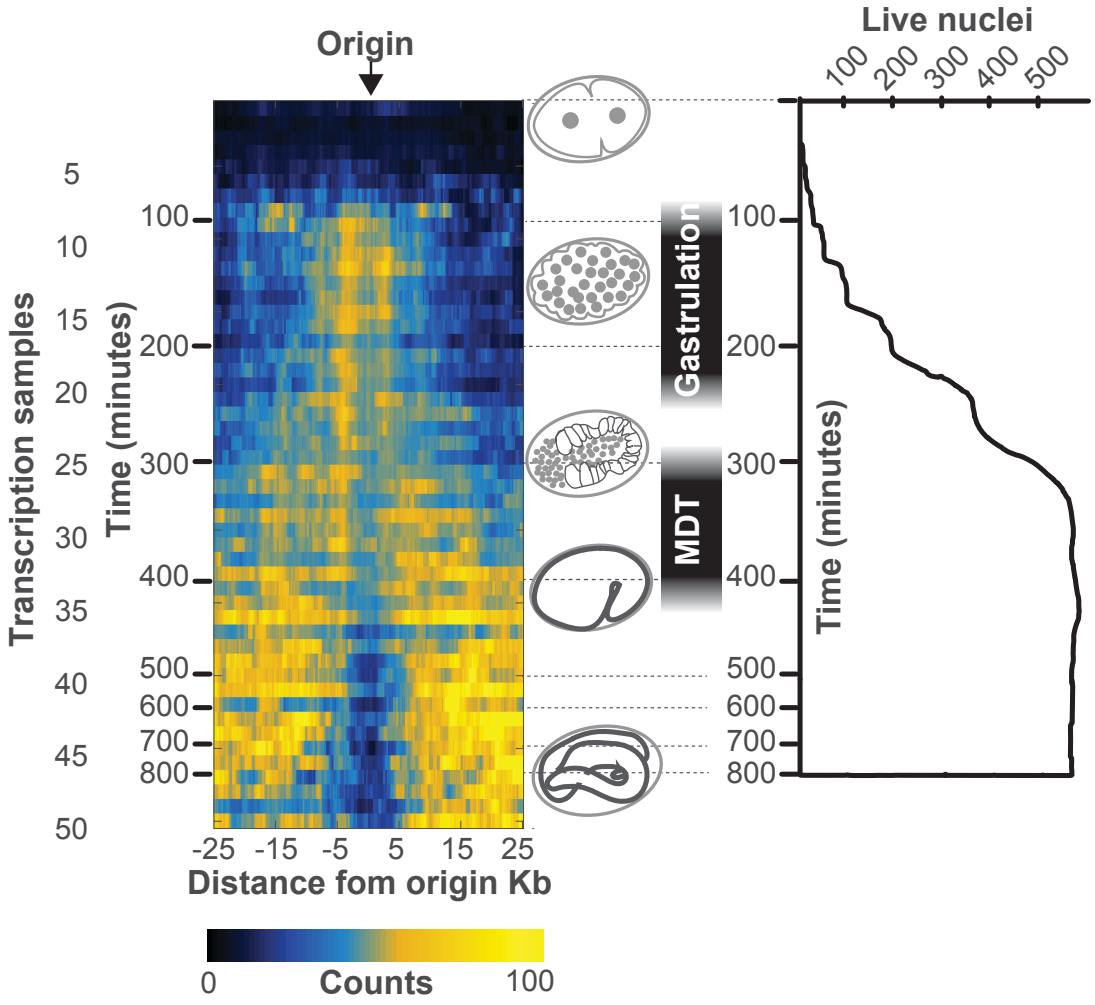

Figure 2.5: Embryonic gene transcription with respect to DNA replication origins. Normalized transcript abundance within ±25 kb of the replication origin was summed for the top 1000 origins for each of the 50 time points through embryogenesis. Data are displayed as a heatmap and arranged according to sample number shown on left, corresponding time in embryogenesis is also indicated. Timing of gastrulation and mid development transition (MDT) calculated by (Levin et al., 2016) are also shown. Live cell nuclei per embryo as calculated by (Sulston et al., 1983) is plotted as a function of time through embryogenesis.

the pattern evolves such that from time point 40 onwards, sites of active transcription now appear anti-correlated with replication origins. The production of zygotic transcripts progressively increases after ventral closure, a milestone embryonic stage which signifies the transition towards a morphogenesis gene expression program (Figure 2.4B) (Hashimony et al 2015).

The drastically different patterns of zygotic gene expression relative to DNA replication origin positions suggested that the positioning of *C. elegans* genes in the genome may influence their embryonic expression timing. To study this further, we performed a gene ontology analysis of genes at varying distances from replication origins. We observe a non-random positioning of genes relative to DNA replication origin positions, with an apparent correlation between genes expressed during proliferative embryonic growth and proximity to DNA replication origins (Figure 2.6). Our analysis revealed that genes whose products are involved growth, embryonic development, cell cycle, gene expression, and chromatin are located less than 5 kb from the nearest replication origins (Figure 2.6). Genes encoding functions involved in tissue-specific developmental processes, such as morphogenesis, neurological system patterning, and hormone mediated signaling, are located > 20 kb away from the nearest DNA replication origin; given our calculated median origin spacing of 40 kb, these genes are located the greatest genomic distance from DNA replication initiation (Figure 2.6).

We identify several spatiotemporal correlations among key genome transactions in the *C. elegans* embryo: DNA replication, chromatin modification, and

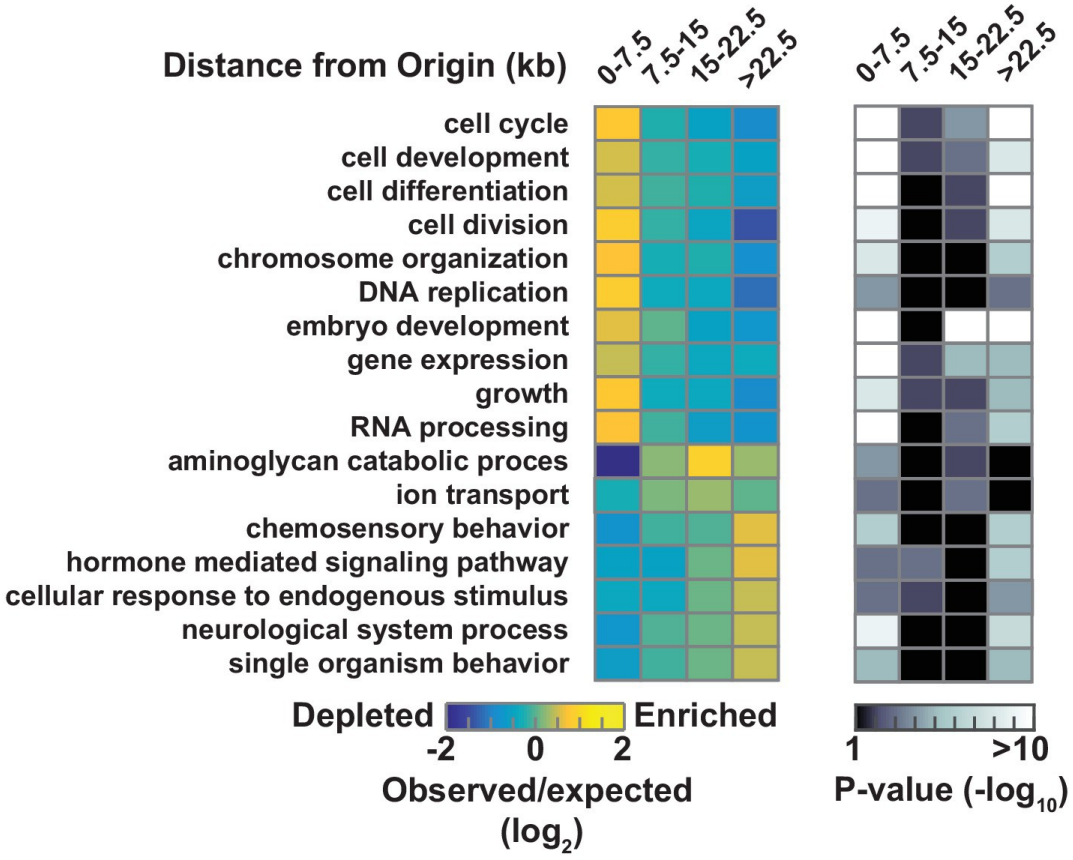

Figure 2.6: Gene ontology analysis for genes at varying distances from the midpoint of a replication origin. Gene ontologies were retrieved from Gene Ontology Consortium (http://geneontology.org/). Left, observed/expected log2 ratio is shown at varying distances from replication origins for each ontology term as a heatmap: color key is below – yellow indicates enrichment, blue is depletion. Right, heatmap showing the calculated p-values (hypergeometric distribution) for the data on left; color key is below.

transcription. Efficient replication origins, represented most in the cell population, are found with 1 kb of histone PTMs associated with transcription, including H3K27ac and H3K4me (Figure 2.3B-C). Replication origins appear to be defined prior to the onset of global zygotic gene activation (Figure 2.4B-C); however, when the genome becomes broadly transcriptionally active, transcription generally occurs in close proximity to the pre-defined origins (Figure 2.5). Approximately 300 minutes into embryogenesis, transcription begins to shift away from the origins, and by 500 min, peak transcription occurs some 15-25 kb from origins (Figure 2.5). Interestingly the alteration in transcription that occurs after 300 minutes is coincident with two developmental milestones: the activation of the morphogenesis transcriptome and the completion of the last wave of embryonic cell divisions (Figure 2.5) (Sulston et al. 1983). This suggests that the activation of replication origins may be coupled with the restructuring of the transcriptome during developmental transitions.

Discussion

Coupling of Early Gene Transcription with Replication Origins

Our finding that replication initiates at specific sites in pre-gastrula embryos (Figure 2.4C) contrasts with the random patterns of initiation observed in the rapid and synchronous cycles of the external developed *Drosophila* and *Xenopus* embryos (Blumenthal et al. 1974; Callan, 1974). However, early embryogenesis in *C. elegans* is markedly different from these organisms: initial cell divisions are asynchronous, and the cell cycle does not dramatically slow during gastrulation (Bao

et al. 2008). Thus, the rapid establishment of the defined replication landscape in *C. elegans* may be an efficient development adaptation to the preformistic mode of lineage establishment, in which cytoplasmic determinants, rather than changes in gene transcription or origin usage, establishes early cell fates (Figure 1.5B). Interestingly, origin proximal genes experience the first wave of zygotic transcription (Figure 2.5) and are continually expressed in replicating cells (Figure 2.6); origin distal regions encode genes whose transcription is delayed until later in development when DNA replication has ceased (Figure 2.6). Our results indicate that the *C. elegans* genome has evolved to couple embryonic transcription with replication; however, such coupling would appear to impose constraints on a system that needs to express a diverse array of genes during embryogenesis and also replicate the 100 megabase genome in 15 minutes.

The *C. elegans* genome organization appears to resolve this issue by clustering genes required for early embryonic cell division near replication origins (Figure 2.6)—importantly, this may allow genes near origins, such as histones, to be replicated at the start of S-phase and have a higher effective copy number than those further away. The spatiotemporal coupling and uncoupling of gene expression with replication may be a general principal underlying the profound transcriptional and morphological changes that occur during embryogenesis across species: genes expressed early are characterized by proteins involved in proliferation, while those expressed late are typically involved in differentiation (Levin et al. 2016).

However, the key question that remains to be addressed is: how is gene transcription remodeled relative to replication origins during the transition from cell proliferation to differentiation? As we observe in Figure 2.5, the genes which were transcribed in the first wave of zygotic transcription no longer appear to be expressed in the late embryo, presumably because genes involved in rapid cell division and growth are no longer required in differentiated cells. One model for consideration might involve a topological remodeling of the chromatin architecture, in the form of chromatin looping or TAD formation, following the final mitotic divisions in late embryos (Dixon et al. 2012). However, analysis of higher order chromatin structure by chromosome conformation capture in *C. elegans* embryos revealed a lack of chromatin structure that would explain a structural remodeling of the embryonic transcriptome on the scale we observe (Crane et al. 2015).

Another important factor we have considered regards the relationship between nascent RNA synthesis and mature mRNA expression in *C. elegans*. Considering the extensive post-transcriptional regulation of the *C. elegans* genome by endogenous RNA pathways and the extensive trans-splicing of pre-mRNA transcripts, it is possible that mRNA transcript levels do not provide an accurate account of RNA polymerase initiation and elongation patterns. The dynamic regulation of gene transcription during distinct phases of embryogenesis will be explored in detail in Chapters 3 and 4 of this thesis.

Transcription-Associated Histone Marks and DNA Replication Origins

We found that DNA replication origins are enriched for histone marks frequently associated with transcription from gene promoters and enhancers, including H3K4 methylation, H3K27 acetylation, and H3K18 acetylation. In agreement with previous studies from *Xenopus* and *Drosophila*, we find that the level of active histone modifications increases according to origin efficiency, suggesting a conserved role for chromatin in specifying replication origins in metazoans (Almouzni et al. 1994; Aggarwal & Calvi, 2004; Lubelsky et al. 2004). The regulatory outcome of H3 acetylation and H3K4 methylation at gene promoters and enhancers is the establishment of local chromatin accessibility, achieved by recruiting transcription coactivator complexes to alter nucleosome density and positioning (Tsukiyama & Wu, 1995; Wysocka et al. 2006). Consequently, the chromatin architecture at regions constitutively modified by H3 acetylation and H3K4 methylation provide stable regulatory domains for the repeated assembly of transcription and replication complexes.

Considering that *C. elegans* histone modification states have been shown to be stable across embryonic and larval stages, an interesting line of inquiry to address is whether the embryonic DNA replication origins we have identified represent common sites of replication initiation in post-embryonic cells divisions, such as those of the endodermal lineage and germline (Evans et al. 2016; Ho et al. 2014). We observe that origin activity remains consistent in proliferating cells of the developing embryo, in spite of differences in gene transcription among the somatic lineages.

These results raise significant questions regarding the relationship between DNA replication and gene transcription during *C. elegans* embryogenesis:

- How do active histone modifications relate to gene transcription in the differentiated cells of the late embryo?
- Are active histone modifications established prior to zygotic transcription, and if so, what mechanism directs their establishment?

Analyzing the interaction between transcription, replication, and chromatin states within the same embryonic time period is critical for developing a unifying model of genome regulation during *C. elegnas* embryogenesis. We identified regions of active histone modifications H3K27ac and H3K4me2, by analyzing ChIP-seq datasets obtained from experiments with a mixed population of embryos, comprising pre-gastrula (8-16 cells) to late gastrula (150-300 cells) (Gerstein et al. 2010). Therefore, the chromatin state in differentiated late embryos, when we observe transcription changes relative to origin positions, is unknown. The following chapter will explore the regulation of active histone modifications and gene transcription during *C. elegans* embryogenesis, with the goal of characterizing the gene regulatory principles guiding embryonic programs of proliferation and differentiation.

Chapter 3

Regulation of the *C. elegans* Embryonic Transcriptome

Introduction

C. elegans Development and Stable Chromatin States

We observe that DNA replication origin efficiency in *C. elegans* embryos is stable, despite the dynamic remodeling of gene transcription relative to origin positions. DNA replication origin efficiency is correlated with the relative enrichment of histone modifications H3K27ac and H3K4me2, marks of active gene transcription. Importantly, the positions of origins and active marks are located within 5 kb upstream of early transcribed genes, consistent with the role of these histone modifications in recruiting transcription complexes. The active modifications are enriched at the promoters of essential genes required for cell viability, encoding functions involved in RNA processing, cell cycle regulation, and transcription. These housekeeping genes are among the earliest to be transcribed during embryogenesis and their expression coincides with replication origin activity. Spatially coupling housekeeping gene transcription to DNA replication origin activity appears to be a common genome organizational principle among metazoans, as TAD boundaries in *Drosophila* and mammalian genomes are defined by clustered sites of replication initiation and constitutively transcribed genes (Hug et al. 2017; Dixon et al. 2013).

Comparisons of the chromatin states between embryos and L3 larvae revealed that histone modification domains are largely stable throughout the lifecycle, segregating into chromosomal domains of germline/broadly expressed and developmentally expressed genes (Evans et al. 2016). The maintenance of histone marks by chromatin modifying enzymes across developmental stages is critical for maintaining transcriptional homeostasis of the germline and the early embryo, evidenced by the sterility and embryonic lethality phenotypes of MES complex mutants (Capowski et al. 1991; Korf et al. 1998; Bender et al. 2006).

Active Histone Marks and Developmental Gene Expression

The marking of gene promoters with H3K27ac and H3K4me2 has been shown to contribute to the differential expression of developmental genes during *Xenopus* and zebrafish embryogenesis (Akkers et al. 2009, Vastenhouw et al. 2010). While the dynamic regulation of histone modifications has been directly addressed in *C. elegans*, the histone acetyltransferase CBP-1 has been shown to play an outsize role in regulating transcription during all stages of worm embryogenesis (Shi & Mello, 1998). CBP-1 activity is required during early embryogenesis to specify several major differentiation pathways, including the mesodermal, endodermal, and hypodermal lineages; inhibition of *cbp-1* expression causes developmental arrest of embryos with no evidence of body morphogenesis and nearly twice the normal complement of embryonic cells (Shi & Mello, 1998). Altogether, this evidence suggests that the transcription of lineage-specific genes during *C. elegans*

embryogenesis is influenced by histone acetylation at developmental gene promoters.

The stability of chromatin states during *C. elegans* development relies on the transmission of histone modifications across generations through the germline (Kelly 2014). Histone methylation is integral to the maintenance of transcriptional balance in the *C. elegans* germline, through the antagonistic activities of chromatin regulators with activating and repressive effects on gene transcription (Gaydos et al. 2012). Deletion of the *C. elegans* H3K4me2 demethylase *spr-5* leads to the inherited accumulation of H3K4me2 at ectopic sites in the germline, leading to a progressive decline in fertility from defects in mitotic germ cell proliferation and gametogenesis (Greer et al. 2014). The transgenerational phenotypes attributed to ectopic H3K4 methylation suggests that the inheritance of modified chromatin states may guide the establishment of developmental gene expression states.

In a study that examined how *C. elegans* histone modifications are established in germ cells and transmitted to offspring, Arico et al. reported that histone modification patterns assembled in the hermaphrodite germline can be retained in the early embryo, despite the chromatin remodeling events which occur at fertilization (Arico et al. 2011). Furthermore, the authors observed that the level of gene expression in adult germ cells correlates with equivalent expression in the somatic lineages of the offspring; using transgenes, it was reported that expression in the parental germline correlated with differential marking of the promoter with H3K4me2 which was replicated and maintained in the early embryo (Arico et al. 2011). Overall,

these results indicate that regulation of chromatin modifications and gene transcription in the *C. elegans* germline may be influential in determining the gene expression state of embryos.

Transcription of Maternal RNAs in the Hermaphrodite Germline

The unique reproductive strategy of *C. elegans* hermaphrodites have profoundly shaped the organization of the genome (Reinke et al. 2004; Reinke et al. 2009; Ortiz et al. 2014). Sequencing of poly(A) selected mRNAs from dissected hermaphrodite gonads revealed that an estimated 10,700 genes are expressed in adult germline, representing approximately half of the *C. elegans* transcriptome; many of these genes encode housekeeping functions that are required for mitotic germ cell proliferation, maternal RNA transcription, meiotic chromosome segregation, and early embryonic DNA replication (Reinke et al. 2004; Ortiz et al. 2014).

C. elegans germ cells are organized with distal-proximal polarity within the syncytium of the hermaphrodite gonad, such that mitotic cells are located at the distal end and transition to meiotic cells in the proximal gonad (Figure 1.3B) (Crittenden et al. 1994). This gradient was recently shown to be organized by spatiotemporal regions of gene expression in the germline; single cell RNA-sequencing revealed that ~6000 genes experience two dynamic shifts in the transcript levels during the exit from mitosis and late pachytene of meiotic prophase I (Tzur et al. 2018). The RNA transcripts expressed during this period are deposited into oocytes and represent the

maternal RNAs that pattern the events of early embryogenesis (Lerner and Goldstein, 1988, Baugh et al. 2003, Gerson-Gurwitz et al. 2016).

The composition and abundance of maternal RNA transcripts is precisely regulated to ensure the provision of gene products that support the early cell divisions of transcriptionally quiescent embryos (Baugh et al. 2003; Gerson-Gurwitz et al. 2016). An analysis of maternal RNA transcript dynamics in staged *C. elegans* embryos revealed significant changes in maternal transcript composition starting at the 4-cell stage and continuing through to gastrulation (Baugh et al. 2003). During this period of time, early cell divisions unequally distribute transcription factors to the founder cells, resulting in the establishment of lineage-specific transcription patterns (Levin et al. 2012).

Transcription Activity in the *C. elegans* Early Embryo

The extensive co-transcriptional and post-transcriptional regulation of RNA transcripts by trans-splicing and small RNA pathways adds significant complexity to the interpretation of steady-state gene expression levels in *C. elegans* (Blumenthal & Spieth, 1996; Billi et al. 2014). The true transcription start sites of many protein-coding genes are unclear because the addition of short, capped spliced leader RNAs to the 5' ends of transcripts occludes the actual location of RNA polymerase initiation (Gu et al. 2012; Chen et al. 2013; Saito et al. 2013). Additional evidence suggests that the landscape of transcriptionally engaged RNA polymerase may not be accurately reflected by steady-state mRNA levels; the development of nascent RNA

sequencing approaches have revealed that promoters are unexpectedly far upstream from the 5' ends of mature mRNA transcripts (Kruesi et al. 2013).

These considerations are relevant to the analysis of zygotic transcription in the early *C. elegans* embryo. The current model of transcription in the *C. elegans* embryo states that broad zygotic transcription initiates at the 28-cell stage, coinciding with gastrulation and the external deposition of embryos outside of the uterus (Baugh et al. 2003; Boeck et al. 2016). Although zygotic transcription becomes more pervasive during gastrulation and a general increase in mRNA transcripts is detected, there is limited empirical evidence defining the number of genes transcribed during early embryogenesis. Bearing in mind the significant overlap in the identity of maternally deposited RNAs and the earliest zygotic transcripts, it is possible that the true extent of zygotic transcription remains unknown due to the inability to discern whether transcripts originated in germline or the early embryo (Gerson-Gurwitz et al. 2016). Indeed there are multiple lines of evidence suggesting that transcription in early embryos is essential for cell fate specification, morphogenesis, and post-embryonic development (Edgar & Wood, 1988; Schauer & Wood, 1990; Powell-Coffman et al. 1996; Shi & Mello, 1998). The extent of pre-gastrula transcription is currently unknown, however assays of RNA polymerase activity in early embryos have provided some clues.

One such study specifically addressed transcription activity by measuring the elongation of previously initiated RNA Pol II transcripts in an assay called nuclear run-on (NRO) labeling—the NRO reaction conditions prevent the initiation of new

RNA Pol II transcription on chromatin and only radiolabels engaged pre-mRNA transcripts which can be isolated and analyzed (Schauer & Wood, 1990). By specifically enriching for pre-gastrula embryo populations (4-8 cells per embryo), the investigators identified high levels of transcription by NRO labeling, estimating that the rate at which nascent RNA is produced in early embryos is 100-300 kb of RNA per nucleus per minute, or approximately 1000-3000 genes (Schauer & Wood, 1990). Furthermore, hybridization of nascent RNA to cDNA libraries revealed that the transcripts encode the essential gene functions required for rapid embryonic divisions, such as DNA replication, transcription, and RNA processing (Schauer & Wood, 1990; Levin et al. 2012). This level of transcription in early embryos is significantly higher than previous estimates and may represent the transcriptional activity required for subsequent stages of embryonic and post-embryonic development (Powell-Coffman et al. 1990; Shi & Mello, 1998).

The pre-determined specification of DNA replication origins in early embryos and the proximity of these origins to gene transcription during gastrulation provided valuable insight into how DNA replication is integrated into the developmental program of *C. elegans* embryogenesis (Figure 2.4B) (Pourkarimi et al. 2016). With the goal of developing a unifying model of gene transcription, DNA replication, and chromatin organization during *C. elegans* embryogenesis, the following section describes results from an assay of dynamic changes in chromatin modifications and gene transcription during a time series of embryo development. By designing a method that enables large populations of embryos to be extracted within a narrow

developmental window, we were able to integrate the datasets of origin activity, mRNA transcription, nascent RNA transcription, and active histone modifications to provide a more comprehensive picture of *C. elegans* genome organization and development.

Results

Culturing Developmentally Synchronized Embryo Populations

The inter-related nature of DNA replication, chromatin, and transcription provides an interesting opportunity to explore how these distinct processes are coordinated during early development. Analyzing the coordination of these processes within the same embryonic time period requires obtaining large populations of synchronized embryos and harvesting material at designated time points spanning embryogenesis. Available ChIP-seq data for embryonic histone modifications H3K27ac and H3K4me2 capture an average chromatin state of early to late stage embryos that might mask punctuated changes occurring in the transcriptional or chromatin landscape during development (Liu et al. 2011, Evans et al. 2016). To address this problem, we developed a large-scale embryo synchronization based on our protocol to harvest pre-gastrula stage *C. elegans* embryos for OF sequencing (Figure 2.4A) (Pourkarimi et al. 2016).

Embryo synchronization for chromatin and transcription analysis was performed with the *egl-30 (tg26)* strain; this line was isolated in a screen for animals with aberrant egg laying timing (EGL) and was found to code for the *C. elegans*

homologue of the heterotrimeric G protein subunit, $G\alpha_q$ (Bastiani et al. 2003). Gravid *egl-30* hermaphrodites, by virtue of their high rate of egg laying, possess fewer embryos within the uterus, effectively narrowing the developmental window during embryo extraction from 4-26 cells down to 2-8 cells (Figure 3.1A) (Pourkarimi et al. 2016). By scaling up the culture of these lines, as many as 250,000-400,000 embryos can be extracted as a synchronous population of 4-8 cell embryos, cultured in vitro during embryonic development, and harvested at select time points for ChIP-sequencing or RNA sequencing analysis. The timing of early cell divisions and the degree of synchrony within bulk embryo populations was confirmed by counting DAPI stained nuclei within the first 100 minutes of development, revealing cell cycle timing comparable to single embryo experiments (Bao et al. 2008, Sulston et al. 1983).

Synchronized Embryos Capture the Maternal RNA and Zygotic Transcriptome

Rigorous validation of this synchronization method requires assessing whether bulk-synchronized embryos can recapitulate embryonic gene regulatory events with precise developmental timing. A series of ten mRNA-sequencing (mRNA-seq) libraries was generated from bulk populations of embryos staged from 4-cells (50 min post-fertilization) to 500-cells (420 min post-fertilization) in order to assess whether the population can capture embryonic transcriptome dynamics. RNA transcripts levels were normalized relative to spike-in controls and were compared to a previously published single-embryo mRNA-sequencing dataset as a reference for embryo gene expression timing (Robinson et al. 2010, Hashimony et al. 2015).

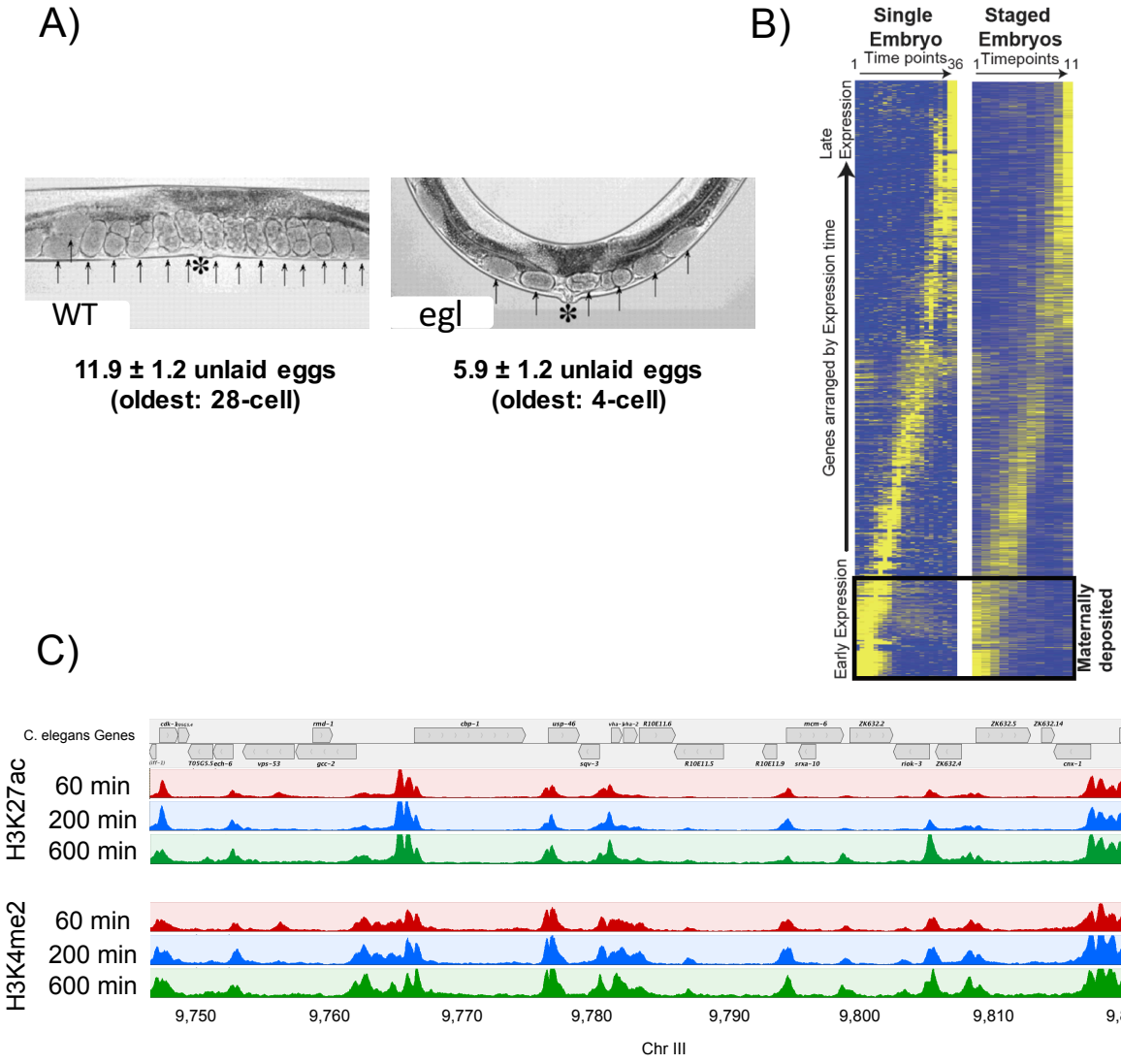

Figure 3.1. Synchronized culture of *C. elegans* embryos. A) Light microscope image of the number of embryos in contained in the uterus of wildtype N2 (left panel) and egl-30 (tg26) (right panel) hermaphrodite worms. Quantitation of embryo numbers from counting n=108 adults. Adapted from Bastiani et al. 2003. B) Validation of developmental synchrony by RNA sequencing. Data generated from single embryo mRNA transcripts isolated at 25 timepoints of embryogenesis up to 420 min post-fertilziation (Hashimony et al. 2015) (left panel) matches 10 time points of mRNA transcription from bulk synchronized embryo populations (right panel); genes are arranged from bottom to top of heatmap by expression timing; dark outlined panel shows maternally deposited RNA transcripts. Gene transcripts normalized relative to spike in controls; mRNA transcripts were aligned to the WS220 genome build and were annotated using Cufflinks software package. C) Genome browser view of normalized H3K27ac and H3K4me2 peaks at early, middle, and late embryonic stages. Genomic range is 70 kb on Chr III; genes arranged on top in grey.

When genes from both datasets were clustered and ranked by expression timing, the synchronized embryo time course up to ~420 min post-fertilization displayed a nearly identical coverage and transcript dynamics to the single embryo RNA-seq dataset, including the maternal and zygotic transcript signature (Figure 3.1B) (Hashimony et al. 2015, Tzur et al. 2018). While the single-embryo mRNA dataset comprised many more time points (n=50) as compared to the bulk synchronized mRNA dataset (n=10), enabling greater time resolution, the overall expression profile of zygotic RNA transcription and maternal RNA decay are remarkably similar (Figure 3.1B). The ranking of both mRNA datasets by gene expression timing reveals a matching pattern of transcript abundance, which importantly confirms that the temporal resolution achieved within the bulk populations accurately represents the timing of early embryogenesis (Figure 3.1B).

Quantitative Comparison of Histone Modifications in *C. elegans* Embryos

Comparisons of the chromatin state between embryos and L3 larvae revealed that histone modification domains are largely stable throughout the lifecycle, segregating into chromosomal domains of germline/broadly expressed and developmentally expressed genes (Evans et al. 2016). The stability of these domains depends on the ability to buffer transcriptional activity through antagonistic regulation by germline chromatin factors (Gaydos et al. 2014)—mutation of histone acetyltransferases or histone methyltransferases has shown to result in a loss of transcriptional homeostasis, embryonic lethality, and progressive sterility (Shi and

Mello, 1998, Katz et al. 2009). However, a quantitative analysis of active histone modifications H3K27ac and H3K4me2 during zygotic genome activation has not been performed in *C. elegans*, as it has been for *Xenopus* and zebrafish embryos (Akkers et al. 2009, Vastenhouw et al. 2010).

I sought to use our embryo synchronization approach to record the dynamics of active histone modifications H3K27ac and H3K4me2 throughout embryogenesis. By purifying large populations of synchronized 4-8 cell embryos, I was able to generate ChIP-sequencing libraries for H3K27ac and H3K4me2 at ten embryonic time points, spanning 60 minutes post-fertilization (4-8 cells) through 800 minutes post-fertilization (~550 cells). Approximately 60,000 embryos per time point were harvested for H3K4me2 and H3K27ac ChIP-seq analysis; all experiments used the identical antibodies generate the reference ModEncode ChIP-seq datasets (Liu et al. 2011, Ho et al. 2014). In order to normalize the enrichment of modifications, I used the ChIP with reference exogenous genome (ChIP-Rx) approach described by the Guenther lab (Orlando et al. 2014). Quantitative comparison of histone modifications relies on the addition of a fixed amount of reference chromatin to each embryo immunoprecipitation; after ChIP, sequencing, and mapping, the *C. elegans* ChIP-seq reads are normalized to the percentage of reference genome reads in the sample (reference adjusted reads per million [RRPM]) (Orlando et al. 2014).

In choosing an organism to use for the ChIP-Rx normalization, we decided upon the fission yeast, *Schizosaccharomyces pombe*. The ChIP-Rx criteria we outlined to achieve an accurate quantitative profile of *C. elegans* histone

modifications were: 1) a reference organism with a similar distribution of the active histone marks we are interrogating in *C. elegans* and 2) an organism with a relatively comparable nucleotide composition as *C. elegans*, which has an unusually high A:T composition (~62%) (Chen et al. 2014). Following the isolation of early stage 4-8 cell embryos, each developmentally staged cohort was harvested and cross-linked with formaldehyde—prior to sonication, an equivalent number of cross-linked *S. pombe* cells were spiked in to each embryonic time point. Thus, immunoprecipitation for active histone marks H3K27ac and H3K4me2 should extract *S. pombe*, as well as *C. elegans*, nucleosomal DNA; sequencing these ChIP libraries provides an internal normalization coefficient based on a consistent percentage of *S. pombe* reads that are amplified in the PCR library. Once the number of *C. elegans* mapped reads is determined (15-50 million reads), it is normalized relative to both: 1) the exogenous genome coefficient from *S. pombe* mapped reads per million (RRPM) and 2) the average number of embryonic cells at each time point.

H3K4me2 and H3K27ac Chromatin State Established in Early Embryos

Mapping of normalized H3K27ac and H3K4me2 reads to the WS235 genome assembly confirmed the characteristic peak profiles over gene promoters, which are concentrated within chromosome centers (Figure 3.1C) (Liu et al. 2011, Gu and Fire, 2010). Peak calling using the Model-based Analysis of ChIP-seq, or MACS, algorithm revealed the presence of 8660 H3K27ac ChIP peaks and 9496 H3K4me2 peaks (Figure 3.2A); surprisingly, many of the H3K27ac and H3K4me2 peaks (active

peaks) are stable in position and intensity for the time spanning embryogenesis (Figure 3.1C), in agreement with previous characterizations of stable domains of broad gene expression (Evans et al. 2016). We also find that active peaks are established prior to gastrulation and thus appear to be independent of zygotic transcription (Figure 3.1C), suggesting that H3K4me2 and H3K27ac modified nucleosomes might be transgenerationally inherited from gamete chromatin as was previously reported in zebrafish and *C. elegans* (Wu et al. 2011, Arico et al. 2011, Greer et al. 2014, Klosin et al. 2017).

The nature of active chromatin marks in pre-gastrula embryos is reminiscent of the site specificity of DNA replication origins that were detected at the same embryonic stage (Figure 2.4C) (Pourkarimi et al. 2016); we decided to identify where active histone marks are positioned relative to DNA replication origins during the embryonic time course. We reasoned that if gene transcription shifted away from DNA replication origin positions in late embryogenesis, our synchronized time course might detect the redistribution of active histone marks to late gene promoters. Mapping the normalized H3K4me2 signal from ten embryonic time points, spanning pre-gastrula embryos to 550-cell late embryos, revealed a striking association between accessible chromatin and DNA replication origins (Figure 3.2). When the normalized H3K4me2 ChIP-seq signal was plotted in 50kb windows around efficient replication origin midpoints, the transcription associated chromatin modification is concentrated within 5 kb of replication origins, even during the period of transcription remodeling at 300-400 minutes into embryogenesis (Figure 3.2: compare mRNA v.

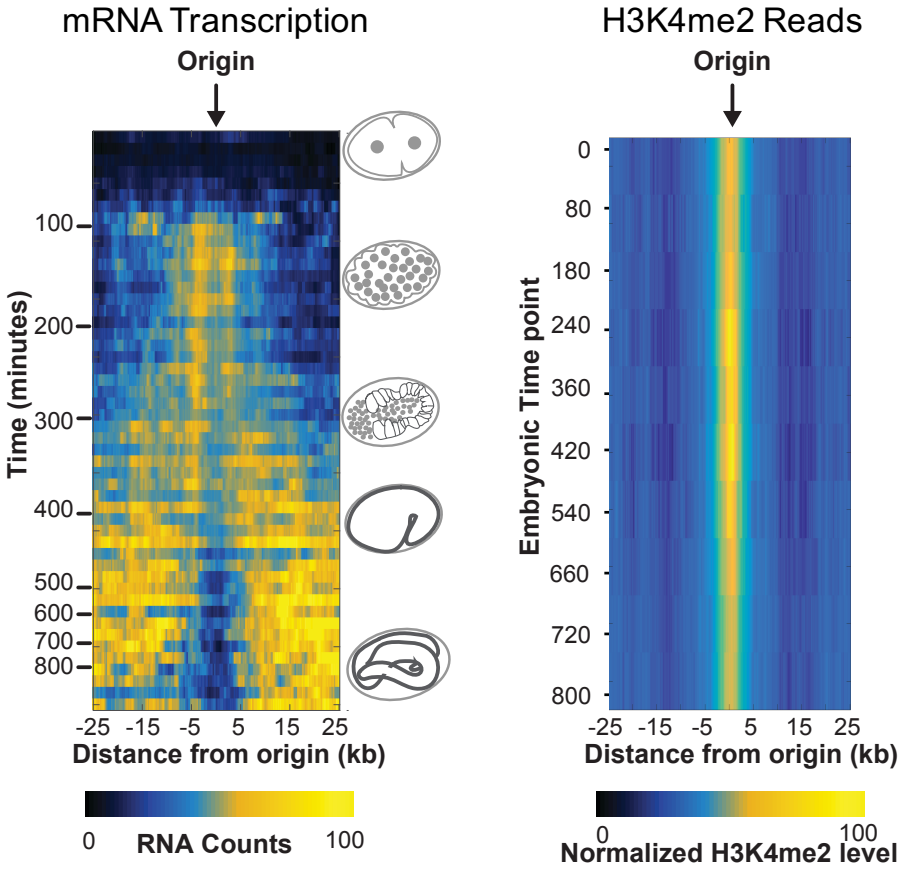

Figure 3.2. Constitutive association of active histone modifications with DNA replication origins. mRNA transcription from the single embryo RNA transcript time series (Hashimony et al. 2015) (left panel) presented as previously in Figure. 2.6. The normalized ChIP-seq signal intensity value corresponding to a total of 6000 H3K4me2 peaks (right panel), called by MACS2 –broad peak calling mode, was binned in 100 bp genomic windows and mapped relative in 25 kb ranges on either side of efficient replication origin midpoints (median efficiency > 0.3). A total of 10 embryonic time points are represented on the Y-axis; normalized RNA counts and H3K4me2 signal intensity is represented by the scale below.

H3K4me2 reads, 300-400 minutes). The profile for H3K27ac ChIP-seq signal relative to DNA replication origins is similar to that of H3K4me2, yet exhibits an apparent decrease in normalized H3K27ac following the completion of mitotic cell divisions (data not shown).

The spatial association between DNA replication origins and active histone modifications in *C. elegans* embryos was noted previously (Figure 2.3), however the dynamic changes in the chromatin state across embryogenesis were not known. Based on this map of H3K4me2 modified chromatin from synchronized embryos, we observe an enrichment of active histone marks with DNA replication origins from the embryonic period spanning 0 minutes to 350 minutes (Figure 3.2; right panel); this spatiotemporal association of active histone marks with origins is consistent with transcription initiation which is occurring at early zygotic gene promoters (Figure 3.2; left panel) (Pourkarimi et al. 2016). However, from the time period of 400-850 minutes, when the embryo is engaged in tissue morphogenesis and organogenesis, an apparent disconnect exists between where the mRNA is transcribed (Figure 3.2; left panel) and where accessible chromatin marks are located (Figure 3.2; right panel). The time course reveals that a majority of the H3K4me2 modification remains associated with DNA replication origin positions (Figure 3.2; right panel), even during embryo morphogenesis, when DNA replication origins are not active and somatic cells are expressing tissue-specific gene transcripts 15-20 kb away from origins (Figure 2.6) (Pourkarimi et al. 2016, Hashimony et al. 2015).

Active Chromatin Marks are Separated into Constitutive or Dynamic Domains

The fact that a large proportion of the active histone marks remain associated with origin positions is surprising given the significant shift in transcription away from origins in late embryos (Figure 2.5) (Pourkarimi et al. 2016). To assess changes in the active chromatin marks, we decided to use the embryonic time series of H3K27ac and H3K4me2 ChIP-seq peaks to identify differentially modified regions during embryogenesis. For each of the histone modifications, we used a differential peak calling algorithm to identify peaks in the pre-gastrula embryo which overlapped by more than 50% with peaks in the late embryo (Neph et al. 2012). This analysis revealed distinct populations of dynamic and constitutive embryonic H3K4me2 and H3K27ac peaks (Figure 3.3A). Among the 8660 total H3K27ac peaks, 4934 peaks (56%) are constitutively maintained from pre-gastrulation (4-8 cells) to late embryogenesis (~400-500 cells) (Figure 3.3B); in the case of H3K4me2 peaks, 6300 out of 9496 total peaks (66%) are maintained (Figure 3.3B). Strikingly, we found a colocalization of H3K4me2 and H3K27ac at 4180 constitutive peak locations, representing nearly 85% of H3K27ac and 66% of H3K4me2 stable peaks called (Figure 3.3B). These peaks did not however appear to have a non-random positioning as was previously observed for DNA replication origins (Figure 2.2B); calculation of the inter-peak distance between constitutively accessible chromatin domains revealed a 22 kb median spacing with an exponential peak distribution (Figure 3.3C).

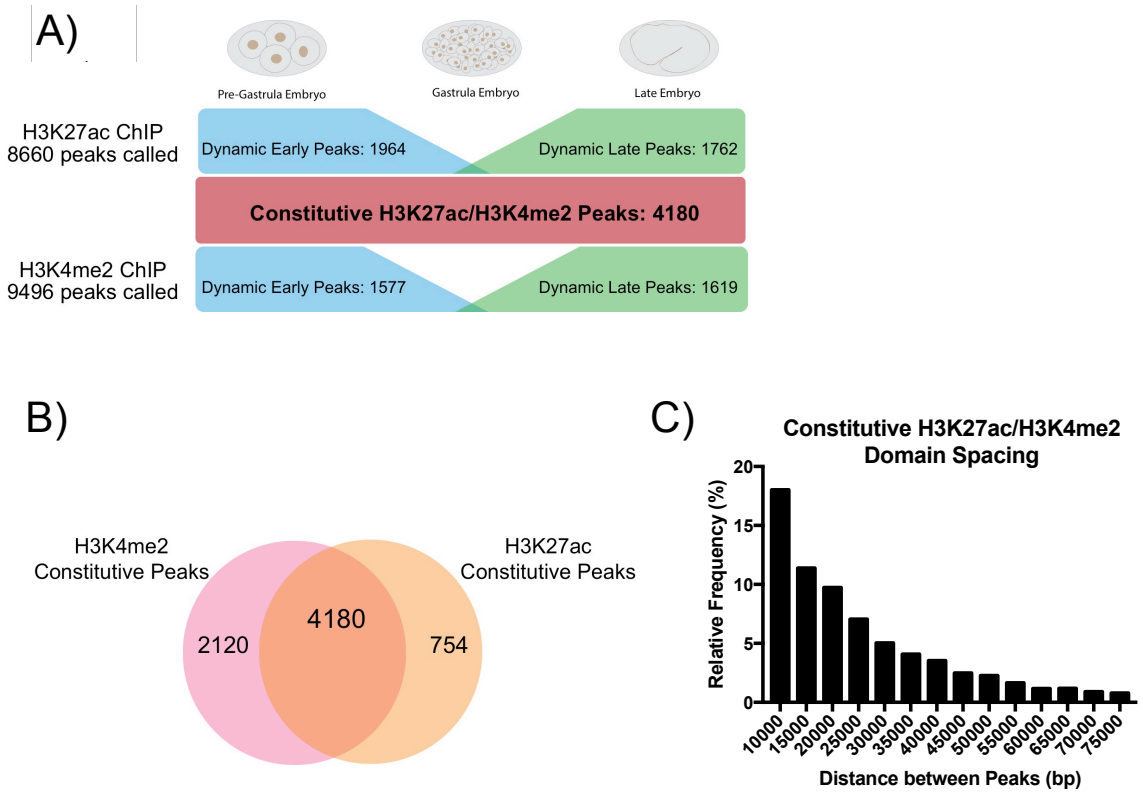

Figure. 3.3. Differential peak calling analysis reveals dynamic and constitutive histone modification domains. A) Categorization of ChIP-seq peaks for H3K27ac and H3K4me2. ChIP-seq reads for each embryonic time point were normalized relative to spike-in reference genome normalization factors and calculated by the reference genome reads per million (RRPM) (Orlando et al. 2014). Peak calling parameters were defined by MACS2 broad peak calling settings. Coordinates from gapped peak files were extracted and used to generate bed files for differential peak calling with Bedops. Constitutive peaks (red) were defined as early (blue) and late (green) peak coordinates with a reciprocal overlap of >50%, while dynamic peaks were those with <50% overlap. Period of embryogenesis represented by cartoon above. B) Constitutive peaks for H3K27ac and H3K4me2 were subsequently called by differential peak analysis to reveal colocalized H3K27ac and H3K4me genomic ranges; overlapping regions in middle of Venn diagram indicate overlapping peaks >50% called by Bedops –overlap option. C) Frequency distribution (%) of constitutive modified accessible chromatin domains. Inter-peak distances are represented on the X-axis and the relative percentage of peaks in each bin are depicted on the Y-axis.

We also identified a population of dynamic H3K27ac and H3K4me2 peaks that were exclusively detected in either early or late embryo populations (Figure 3.3A). Peaks lost in the transition from pre-gastrula embryos to late embryos were defined as “dynamic early peaks”, while peaks acquired only in late embryos were defined as “dynamic late peaks” (Figure 3.3A). In contrast to the high percentage of constitutive peaks marked by colocalization of H3K4me2 and H3K27ac (Figure 3.3B), the dynamic population of peaks had far fewer colocalized peaks, suggestive of cell-type specific, rather than broad patterns of gene expression (Heintzman et al. 2009).

Altogether, these findings suggest that the landscape of active histone modifications in *C. elegans* embryos comprises regions constitutively marked by H3K27ac and H3K4me2 and regions of dynamic modification that are correlated with the temporal expression of developmentally regulated genes.

Constitutive Peaks Correlate with DNA Replication Origin Efficiency

The association between histone modifications and the specification of DNA replication origins has been reported in several metazoan species and may play a role in maintaining DNA replication timing programs in animals (Lubelsky et al. 2015; Pourkarimi et al. 2016). To determine the relationship between DNA replication origin efficiency and the active histone modification landscape, we calculated DNA replication origin efficiencies (n=1800, pre-gastrula origin efficiencies >0.3) in 100 bp windows across the genome and plotted the mean origin efficiency relative to the midpoint of constitutive peaks (n=4180, overlapping H3K4me2/H3K27ac), dynamic

early peaks (n=3541, combined H3K4me2 and H3K27ac), and dynamic late peaks (n=3381, combined H3K4me2 and H3K27ac) (Pourkarimi et al. 2016). As is shown in Figure 3.4A, origin efficiencies display a “bell-shaped” distribution relative to constitutive peaks, suggesting that the efficiency of a DNA replication origin, or how frequently the origin fires in the cell population, may be influenced by regulation within these particular chromatin domains. However, we find that not all accessible chromatin states are associated with DNA replication initiation, as the dynamic peaks in early and late embryos do not exhibit any discernible correlation with replication origin efficiency (Figure 3.4A). Based on their transient appearance in the embryo, the dynamic chromatin states may be regulated separately from DNA replication and instead might specifically regulate lineage-specific gene transcription in specialized cell types, rather than broadly expressed genes (Figure 3.2A).

The distribution of these accessible chromatin states relative to DNA replication origins suggests that the constitutive maintenance of these domains may be important for the developmental demarcation of origin positions (Figure 3.4A). The constitutive peaks genome-wide exhibit a median spacing of ~20kb (Figure 3.3C), situating an average of two accessible chromatin domains between adjacent replication origins (Pourkarimi et al. 2016). These results suggest that the maintenance of these accessible chromatin domains in *C. elegans* may help reinforce consistent transcription and replication initiation across different developmental stages and cell types (Schwaiger et al. 2009, Pope et al. 2014, Lubelsky et al. 2015, Petryk et al. 2016).

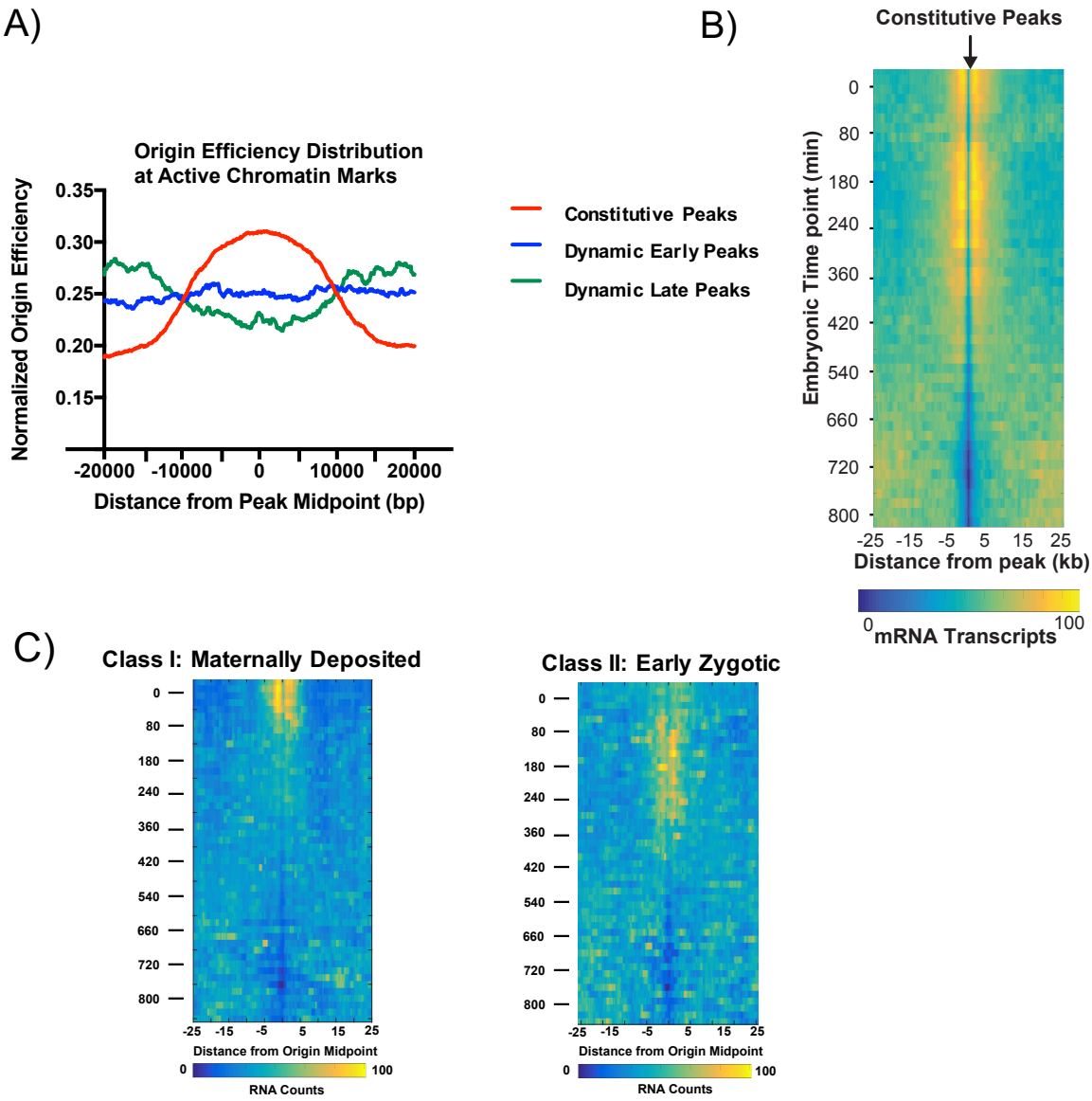

Figure. 3.4. Constitutive domains of active histone modifications are associated with DNA replication origins and the early embryonic transcriptome. A) The median DNA replication origin efficiency was binned in 100 bp windows throughout the genome; those values were plotted in 40 kb genomic ranges to either side of active histone modification peak midpoints. DNA replication origin efficiency datasets from Pourkarimi et al. 2016; efficient origins >0.3. B) Constitutive peaks associated with early embryonic transcriptome. Hashimony et al. 2015 normalized mRNA time series was mapped as in Figure 2.6. Midpoints of constitutive peaks (n=4180) were mapped in the center of the heatmap, mRNA transcript values were normalized such that each of the 50 time points had a sum of gene expression equal to 1. Scale of mRNA transcript level shown on bottom. C) Maternal RNA transcripts map to constitutively accessible chromatin regions associated with replication origins. Class I (growth genes) and Class II (lineage-specifying transcription factors) maternal transcripts (n=3900), specified by Baugh et al. 2003, were respectively mapped ±25 kb from the midpoints of constitutive active mark peaks (n=4180).

Constitutive Peaks Associated with the Early Embryonic Transcriptome

During gametogenesis, mitotically cycling stem cells exit mitosis, accumulate proteins required for entry into meiosis, and proceed through two consecutive rounds of meiotic cell divisions (meiosis I and meiosis II) (Kimble and White, 1981, Seydoux and Schedl, 2001, Kadyk and Kimble, 1998). The spatiotemporal expression of the *fem* and *fog* sex determination genes in the *C. elegans* hermaphrodite germline restricts spermatogenesis to the L4 larval stage before switching to oogenesis for the remainder of adulthood (Hirsch et al. 1976, Hodgkin and Brenner, 1977, Schedl and Kimble, 1988, Hodgkin 1986). During oocyte development and maturation, maternal RNAs accumulate within the egg cytoplasm and are subject to post-transcriptional regulation by translational control factors, RNA binding proteins, and RNAi pathways—the interaction of these pathways regulate gene expression in the early embryonic divisions (Reinke et al. 2000, Merritt et al. 2008, Seth et al. 2013, Gerson-Gurwitz et al. 2016). We have previously determined that DNA replication origin positions are established independent of zygotic transcription (Figure 2.4B-D); the map of the early embryo H3K4me2 state (Figure 3.2; right panel, 0-100 min) reveals that active marks are deposited at the eventual sites of zygotic transcription initiation prior to when the genes are expressed during gastrulation (Figure 3.2; left panel; 100-200 min).

While little is known regarding RNA polymerase activity in germ cells, our map of active histone marks from early embryos provides some clues as to how the early embryo chromatin state may be established. The early *C. elegans* embryo contains

maternal RNA transcripts that largely overlap with early zygotic expressed genes (Baugh et al. 2003)— similarities in the *C. elegans* germline and embryonic transciptomes indicate that the constitutively accessible chromatin state might be inherited from the germline to the early embryo to regulate zygotic gene transcription (Figure 3.2; right panel) (Ghosh and Seydoux, 2008, Tzur et al. 2018, Gerson-Gurwitz et al. 2016).

To investigate this possibility, we compared constitutive peak positions (n=4180) with the whole embryo transcriptome time series derived from individual embryos representing the one-cell stage through hatching (Hashimony et al. 2015, Pourkarimi et al. 2016). We observe a tight clustering of maternal RNA transcripts around constitutive peak midpoints in the first 60 minutes of embryogenesis, the period prior to broad zygotic transcription (Figure 3.4B; 0-60 min). There is a slight decrease in the level of these maternal RNA transcripts at 80 minutes, possibly reflecting the period of maternal RNA clearance and the first wave of zygotic transcription (Baugh et al. 2003) (Figure 3.4B; 80 minutes), followed by a burst of gene transcription near constitutive peaks from gastrulation (100 minutes) until the beginning of morphogenesis (~350 minutes) (Figure 3.4B; 100-350 minutes) (Levin et al. 2012, Pourkarimi et al. 2016).

The relationship between the embryonic transcriptome and constitutive peaks breaks down after the final cell divisions; this decoupling of gene expression from constitutive peaks at ~350 minutes post-fertilization likely explains the reported shift in transcription relative to DNA replication origins after their inactivation in late

embryos (Pourkarimi et al. 2016). Interestingly though, we were able to identify a subpopulation of chromatin peaks which appear associated with both DNA replication origins and the proliferative gene expression program during the early stage of embryogenesis (Figure 3.4A-B). Altogether, our data suggest that the transcription of maternal RNAs in the germline and the first wave of zygotic genes in the embryo are correlated with a subset of regulatory domains that are constitutively marked by H3K27ac and H3K4me2 and are clustered in close proximity to DNA replication origins.

Analyzing Nascent RNA Transcription in Early and Late Embryos

The transcription of essential housekeeping genes proximal to DNA replication origins and active histone marks characterizes the pattern of gene transcription during early embryogenesis (Levin et al. 2012, Pourkarimi et al. 2016). However the association of gene transcription with these regions peaks is lost during the transition to the morphogenesis program, when gene transcription appears to occur 15-25 kb away from domains of early gene expression (Figure 3B) (Hashimony et al. 2012, Pourkarimi et al. 2016).

Considering the extensive post-transcriptional regulation of mRNA levels through RNA degradation pathways and RNAi mediated gene silencing, analysis of mature mRNA transcripts might not provide an accurate recording of nascent RNA polymerase transcription genome-wide (Core et al. 2008, Kaikkonen et al. 2013). To compare the patterns of RNA polymerase elongation at distinct developmental times,

we decided to analyze published global run-on sequencing (Gro-seq) libraries from gastrulation stage embryos and starved L1 larvae (Kruesi et al. 2013). In an adaptation of the nuclear-run on (NRO) reaction, Gro-seq libraries are generated by pulse labeling isolated nuclei with affinity-tagged nucleotides, such as brominated UTP, under conditions that ensure only transcripts engaged in transcription with RNA Pol II will be labeled (see Methods) (Kruesi et al. 2013, Jonkers et al. 2014). When the total RNA is purified and fragmented, nascent transcripts can be enriched by immunoprecipitation to provide a read-out of RNA polymerase elongation (Kruesi et al. 2013). Because nascent RNA transcripts are enriched above mature mRNA transcripts, this technique might enable us to develop a mechanistic understanding of the transcriptome remodeling between early and late embryogenesis

Transition to Larval Development Triggers Pervasive Transcription Elongation

To provide a stable genomic landmark from which to compare developmental changes RNA polymerase elongation patterns, we decided to map embryo and L1 Gro-seq reads relative to the constitutive H3K27ac peaks we identified in the differential peak analysis (Figure 3.5). To determine the positional relationship between nascent RNA synthesis and mature mRNA production, we ranked genes on the Watson strand by the distance between the 3' gene end and the midpoint of the nearest constitutive H3K27ac peaks. We then plotted the normalized mRNA (Hashimony et al. 2015) or nascent (Kruesi et al. 2013) RNA signal from embryos

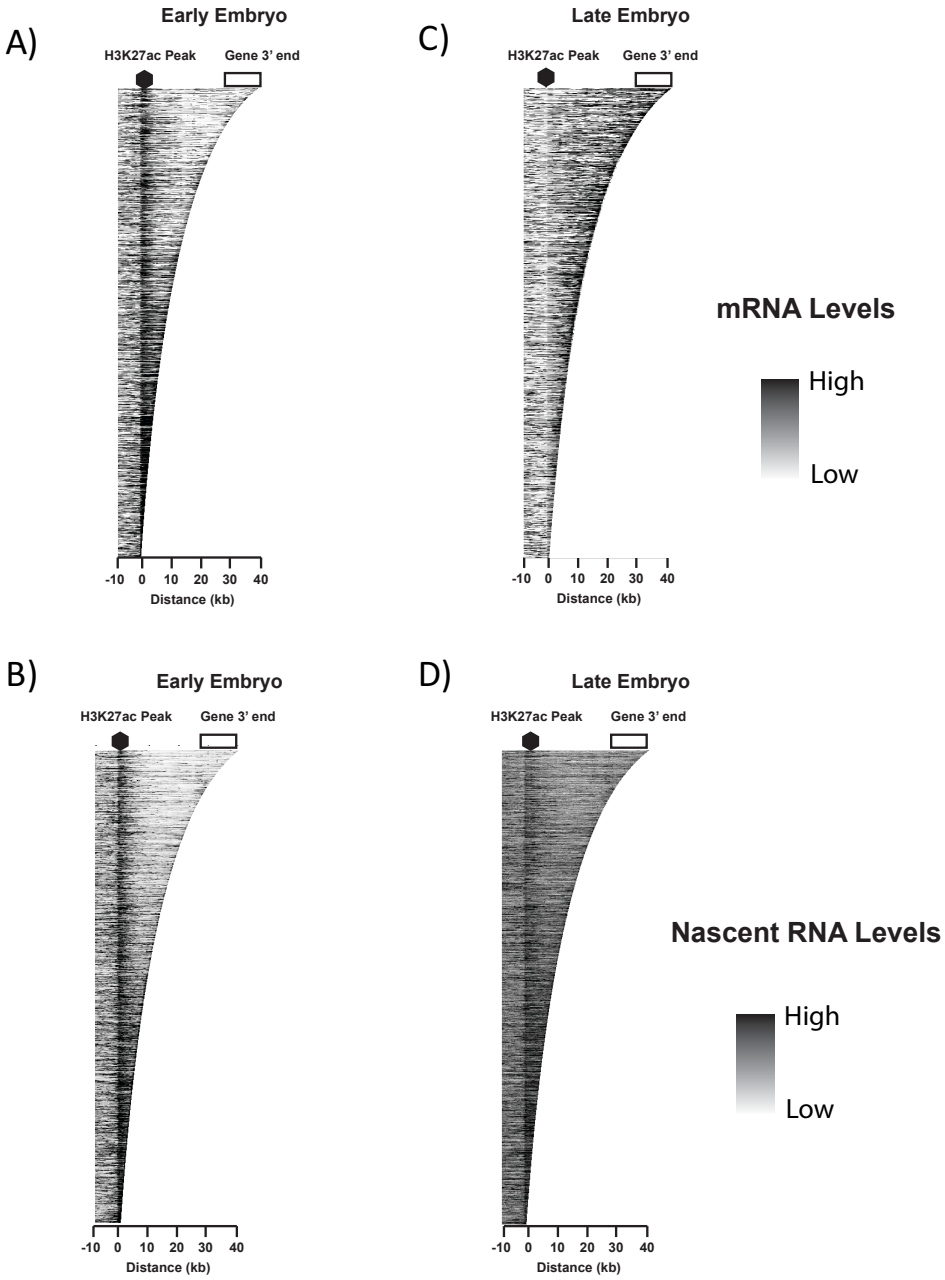

Figure 3.5. Remodeling of RNA polymerase elongation coupled to transition from embryo to L1 larvae. The 3' end of Watson strand genes were plotted within 50 kb genomic range (-10 kb upstream, +40 kb downstream) relative to the closest upstream constitutive H3K27ac peak midpoints; genes were ranked according to distance, with the genes closest to an H3K27ac peak on the bottom and the genes farthest from a peak on the top of the heatmap. A) 200 minute stage embryo mRNA transcript levels and (Hashimony et al. 2015) and B) Gro-Seq signal (Kruesi et al. 2013) were compared to 800 minute stage embryo mRNA transcripts and D) L1 Gro-Seq signal (Kruesi et al 2013). Scale for normalized mRNA and nascent RNA levels shown on right.

and larvae in 50 kb genomic ranges, encompassing a 10 kb upstream region and a 40 kb downstream region with respect to H3K27ac peaks.

The results presented in Figure 3.5 illustrate distinctive patterns of nascent RNA transcription and mRNA production across two different *C. elegans* stages, embryos and L1 larval development. The embryo mRNA and nascent RNA data are representative of gastrulation stage embryos—plotting mRNA transcripts relative to H3K27ac peaks reveals that most mRNA transcripts accumulate within 10 kb from the nearest accessible chromatin region (Figure 3.5A). The nascent RNA transcript profile is similar to the mRNA transcript profile, as initiation and elongation of RNA polymerase in gastrulation stage embryos is largely confined to <10 kb transcription units downstream of the nearest H3K27ac peak (Figure 3.5B).

Starved L1 larvae exhibit a gene expression profile that is largely analogous to late stage embryos (Boeck et al. 2016); we reasoned that comparing the transcriptomes of L1 larvae and gastrulation stage embryos may explain the remodeling of gene transcription during morphogenesis (Figure 2.5). We find that in contrast to the embryo mRNA transcript profile (Figure 3.5A), the L1 mRNA transcripts accumulate 20-25 kb away from the nearest active histone modification (Figure 3.5C). However the nascent RNA transcripts from L1 larvae reveal a remarkable degree of transcription elongation relative to active histone marks (Figure 3.5D). Contrasting with the confined transcription elongation during the gastrulation stage of embryogenesis (Figure 3.5B), the starved L1 transcriptome exhibits a dramatically remodeled transcription landscape evidenced by nascent RNA

sequencing reads extending from the nearest H3K27ac peak to 30-40 kb downstream (Figure 3.5D).

Embryonic Patterns of Transcription Initiation and Elongation

Interestingly, we observed that while mRNA signal proximal to H3K27ac peaks disappears in the transition from embryo to L1, the nascent RNA signal suggests that transcription is still occurring at regions. To investigate whether the transcriptional remodeling pattern occurs during the transition to morphogenesis in late embryos, we applied our system to synchronize *C. elegans* embryos to generate a time series of nascent RNA transcription. We generated Gro-seq libraries from pre-gastrula embryos, mixed early embryos, and late embryos using the previously described approach (Kruesi et al. 2013). We mapped the nascent RNA reads to the WS235 genome and observed a high level of RNA polymerase iniitation from H3K27ac peaks at the 5' end of early expressed genes (Figure 3.6). However, as embryogenesis progressed, we observed a gradual extension of RNA polymerase elongation from the early expressed gene to tens of kilobases downstream (Figure 3.6; compare Gro-seq signal from 60, 200, and 600 min track). To observe whether this shift in RNA polymerase elongation is occurring across the genome, we generated a meta-gene plot that displays the distribution of Gro-seq signal in a 30 kb region, extending 5 kb upstream to 25 kb downstream of the constitutive H3K27ac and H3K4me2 peaks (n=4180) (Figure 3.7). The normalized mRNA signal reveals a

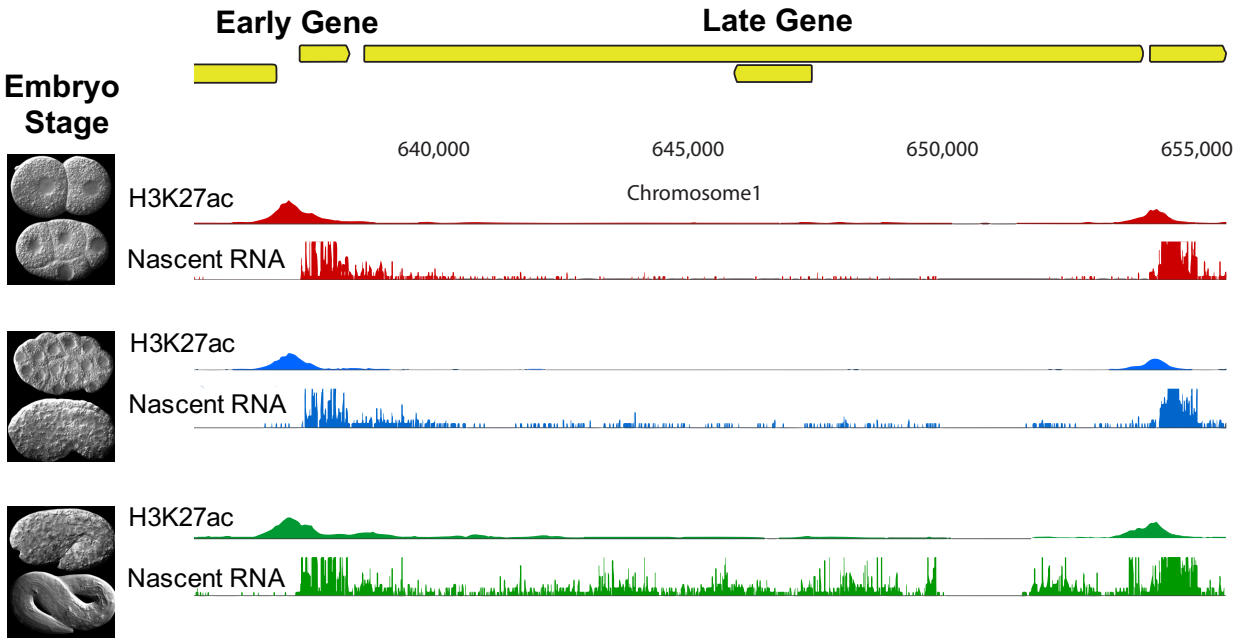

Figure 3.6. Gradual RNA polymerase elongation corresponds to transcription remodeling during *C. elegans* morphogenesis. Representative genome browser view of 60 min pre-gastrula embryo (red), 200 min gastrulation stage embryo (blue), and 600 min late stage embryo (green) histone modifications (top row) and nascent RNA sequencing reads (bottom row). Early transcribed gene and late transcribed genes from the Watson strand on chromosome I (yellow) serve as positional landmarks to demonstrate gradual elongation of RNA polymerase in the transition from early proliferative growth (60min; red) to late morphogenesis transcription (600min; green).

familiar pattern, in which transcripts in the 60 min and 200 min time points are highest proximal to the accessible chromatin peaks, while transcripts begin to shift 10-20 kb downstream in the 600 min embryo (Figure 2.5; Figure 3.7A). Surprisingly though, we observe in the late embryo that RNA polymerase initiates proximal to accessible chromatin peaks and elongates 10-20 kb downstream to express distal somatic genes involved in cell differentiation and organogenesis (Figure 3.7B; Figure 2.6). Contrary to expectation, this pervasive elongation in late embryos occurs in the absence of global remodeling of the chromatin landscape (Figure 3.2), and in the absence of higher order chromatin structures such as TADs or enhancer loops (Crane et al. 2015). Additionally, the pattern of RNA polymerase initiation in late embryos (Figure 3.6; green track) suggests that growth genes which are only expressed in differentiating cells are still being transcribed (Figure 3.7B). Combined with the Gro-seq data from recently hatched L1 worms (Figure 3.5C-D), our analysis of nascent RNA transcription suggests that the switch from cell proliferation to animal morphogenesis is coupled to a change in RNA polymerase elongation (Figure 3.6-3.7). Starting from the final cell divisions at 350 min post-fertilization, RNA polymerase begins extending downstream of early expressed genes and travels tens of kilobases to somatic differentiation genes (Figure 3.5).

Collectively, this data suggests that the dynamic remodeling of the transcriptome that is observed in the transition from early to late embryogenesis may be indicative of a genome wide transition from the germline/embryonic transcriptional program to the somatic transcriptional program.

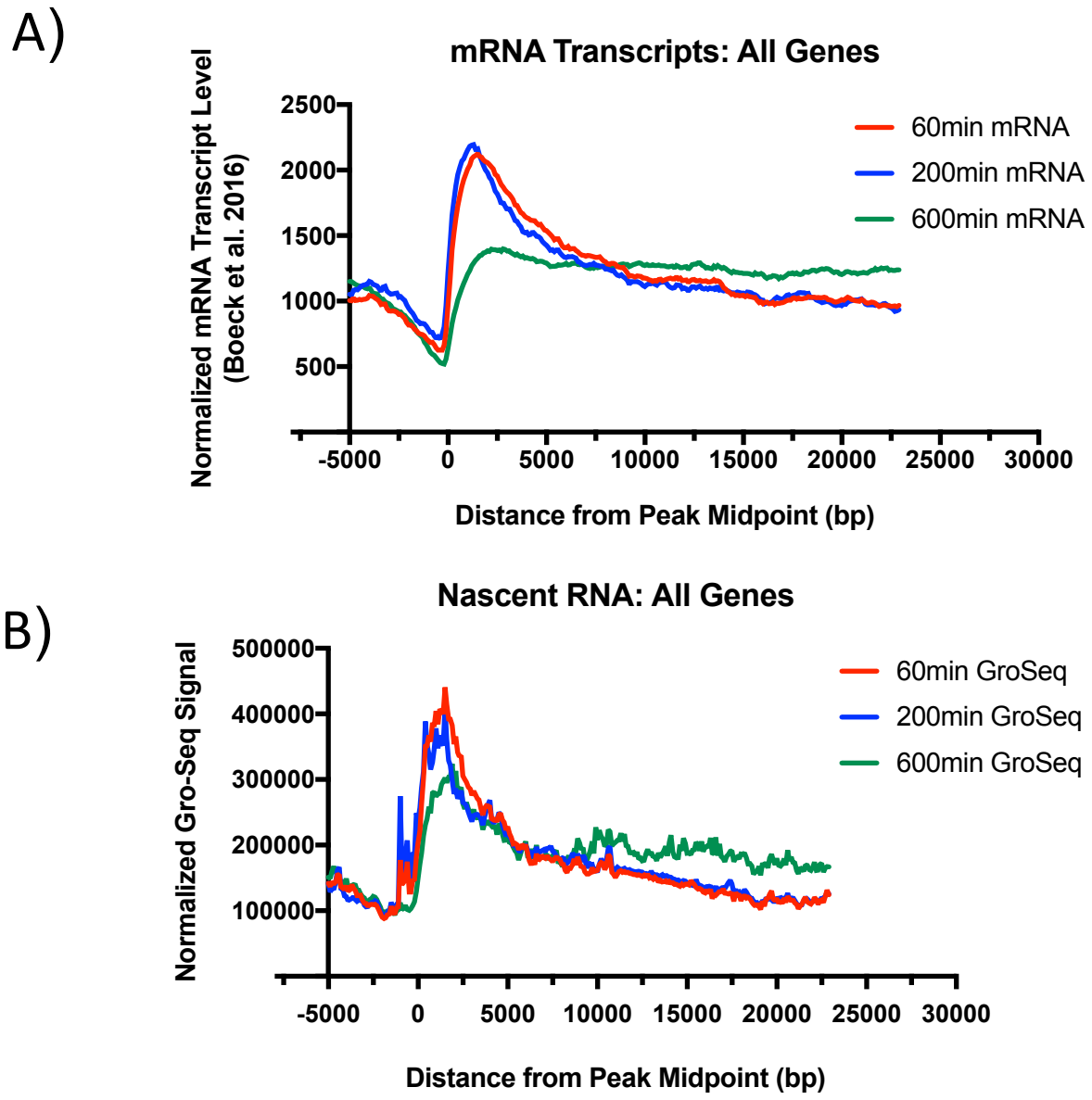

Figure 3.7. Meta-analysis of gradual transcription remodeling during late stages of embryogenesis. A) Normalized mRNA transcript levels (Hashimony et al. 2015) from indicated embryonic time points were summed in 100 bp genomic windows and plotted relative to origin-associated constitutive peaks (n=4180) in 25 kb genomic ranges (-5 kb upstream and +20 kb downstream) relative to the peak midpoint B) Normalized Gro-Seq reads from indicated embryonic timepoints are plotted relative to constitutve peak midpoints; Gro-seq libraries were generated as previously reported (Kruesi et al. 2013); Watson and Crick Gro-seq reads were mapped separately using Samtools flags; base scores of normalized Gro-seq signal from respective strands were summed in 100 bp genomic windows and plotted relative to peak midpoints.

Discussion

Landscape of Active Histone Modifications in *C. elegans* Embryos

In order to analyze the temporal establishment of active chromatin marks in the *C. elegans* embryo, we developed an *in vitro* culturing system that enabled the isolation and assessment of highly synchronized embryos at time points spanning embryonic development. This technique provided enough material to perform ChIP-sequencing of active histone modifications from 4-8 cell embryos, a critical developmental stage that has eluded experimentation because the first six embryonic divisions occur in the uterus of the adult hermaphrodite. Applying ChIP-sequencing to developmentally staged embryo populations, we identified a total of 8,000 H3K27ac peaks and 9,000 H3K4me2 peaks across embryogenesis (Figure 3.1); by virtue of our synchronized time course, we were able to categorize these peaks into dynamic and constitutive populations by comparison of early and late embryo chromatin states.

The identification of differentially modified H3K27ac and H3K4me2 domains provides a valuable resource to analyze developmental gene regulation at distinct stages of *C. elegans* embryogenesis. The fact that the dynamic peaks of active histone marks exclusively correlates to early or late embryogenesis suggests that: 1) these peaks are correlated with transcription in only a subpopulation of tissues, or 2) these domains are reversibly modified by chromatin regulators at specific gene promoters. While are unable to determine from which cell type the modified nucleosomes originated, the dynamic nature of H3K27ac modifications in other

metazoan embryos suggests that the temporal binding of transcription factors may relate to the transience of the histone modifications (Bogdanovic et al. 2012). A key line of inquiry to pursue relates to how dynamic histone modifications correlate with the recruitment of conserved transcription factors and the activation of their target genes at distinct embryonic stages.

Transcription factor binding at developmentally regulated gene promoters is known to facilitate the recruitment of numerous chromatin regulatory enzymes which can differentially modify nucleosomes at the promoter (Koche et al. 2011). A complex comprising the zinc-finger transcription factor MEP-1 and two *C. elegans* homologues of the mammalian nucleosome remodeling and deacetylase (NuRD) complex, LET-418/Mi-2 and HDA-1/HDAC-1, are reported to control the levels of histone acetylation at dynamically regulated germ and somatic gene promoters (Unhavaithaya et al. 2002, Anderson et al. 2006). The disappearance of histone acetylation in dynamic early peaks may be regulated by the NuRD complex, as the appearance of the peaks in early embryos suggests the marks may be inherited as a result of germline transcription (Unhavaithaya et al. 2002; Arico et al. 2011). Distinguishing these populations of dynamic peaks provides a valuable resource to analyze developmentally regulated genes during *C. elegans* embryogenesis.

Constitutive Domains and DNA Replication Origin Efficiency

We also found that the chromatin domains with the highest level of H3K27ac and H3K4me2 marks were established prior to zygotic transcription, much in the

same manner as DNA replication origins (Figure 2.4; Figure 2.6) (Pourkarimi et al. 2016). By focusing our analysis on these constitutive chromatin domains, we found that these peaks are: stably associated with DNA replication origin positions for the all of embryogenesis (Figure 3.3), directly correlated with DNA replication origin efficiency (Figure 3.4), and characterized by the colocalization of H3K4me2 and H3K27ac modifications (Figure 3.1: Figure 3.2). These results explain many of our previously reported associations between active histone marks and DNA replication origins, particularly that H3K4me2 and H3K27ac scale with origin efficiency (Pourkarimi et al. 2016). We find that the efficiency of DNA replication origins exhibits a “bell-shaped” distribution relative to their distance from constitutive peak midpoints; by comparison, origin efficiency appears uncorrelated with the dynamic early peaks and anti-correlated with the dynamic late peaks (Figure 3.4). The relationship between origin efficiency and these constitutive domains affords the opportunity to address several questions relating to the establishment of the embryonic DNA replication origins:

- What is the nucleosome positioning at these regulatory domains and how is this architecture maintained in the face of DNA replication?
- How are replication complexes assembled at these sites and to what extent is their assembly dependent upon histone modifications and transcription?
- Do these domains represent the sites of DNA replication initiation in post-embryonic cell divisions?

Addressing these questions may reveal significant insights as to how DNA replication origins are specified in embryonic and post-embryonic developmental stages and whether chromatin modification has a role in stably transmitting DNA replication origins from generation to generation.

Constitutive Domains and Embryonic Transcription

By comparing the overlap between constitutive H3K27ac peaks and H3K4me2 peaks, we found that approximately ~4000 domains exhibit a 50% reciprocal overlap in these active histone modifications (Figure 3.2). The colocalization of H3K27ac and H3K4me2 at these ~4000 constitutive peaks is similar to a chromatin signature in *Drosophila* and mammalian genomes that is associated with a high occupancy of transcription factors and coactivators involved in constitutive cycles of RNA polymerase initiation (Moorman et al. 2006, Bonn et al. 2012, Visel et al. 2009). Furthermore, the maintenance of the constitutive peaks is consistent with previous characterizations of H3K4me2 and H3K27ac enrichment across the *C. elegans* genome as being associated with the promoters of genes expressed broadly across cell types (Figure 1.7) (Ho et al. 2014; Evans et al. 2016). It is not known whether these histone marks co-occur on the same nucleosomes, are maintained by shared chromatin regulatory pathways, or whether the modifications are unrelated and simply correlate with local chromatin accessibility. Addressing these questions might yield interesting insights into the possible co-regulation of active histone modification domains.

The sustained marking of these domains with active histone modifications is reminiscent of the canonical promoter architecture at constitutively transcribed genes, which exhibit a clustering of transcription factor binding sites at accessible, nucleosome depleted regulatory domains (Segal et al. 2008, Lee et al. 2004). Guided by our previous finding that efficient DNA replication origins lie in close proximity to housekeeping genes, we decided to map early embryonic mRNA transcripts relative to the constitutive peak positions (Figure 3.4) (Pourkarimi et al. 2016). We observed a dramatic clustering of early embryonic transcripts spanning the time period from fertilization to the beginning of morphogenesis, at which point the association breaks down and becomes uncoupled from constitutive peak positions (Figure 3.4).

The recent publication of a spatiotemporal gene expression map of the syncytial *C. elegans* germline has provided valuable insight into the dynamics of maternal RNA transcription (Tzur et al. 2018). The authors observed that over 6000 genes exhibited restricted spatiotemporal expression in accordance with known progression of germ cell development; the greatest changes in gene expression occur during late pachytene of meiotic prophase I, prior to the general reduction in transcription noted to occur in the following diakinesis phase (Tzur et al. 2018; Kelly et al. 2002; Gilbert et al. 1984). These mRNA transcripts are subsequently taken up by oocytes as they mature, providing the gene products required to complete the meiotic divisions and early embryogenesis (Lerner & Goldstein, 1988; Kelly et al. 2002). Our results suggest that the constitutive maintenance of active histone

modifications at the promoters of germline expressed genes may mediate the transfer from maternal to zygotic gene regulatory control (Figure 3.4).

There are several possible explanations to explain the association of maternal RNA transcripts with the constitutive peaks. One possibility is that the reproductive mode of *C. elegans* hermaphrodites led to the adoption of a common chromatin state that is compatible with the transcriptome of germ cells and early embryos. This mechanism might minimize time spent in the vulnerable embryonic period by avoiding a timely remodeling of the transcriptional state between gametogenesis and early zygotic transcription. Another possibility might relate to the positioning of essential housekeeping genes relative to DNA replication origins (Figure 2.5); the close proximity of cell viability genes to DNA replication origins may suggest that evolution of the *C. elegans* genome was driven to produce the optimal chromatin state for the rapid expression of genes essential to germ cell development and early embryogenesis. A more extensive discussion of this model will be presented in Chapter 4.

Transcriptome Remodeling in Late Embryos

The publication of nascent RNA sequencing datasets allowed us to investigate the relationship between patterns of transcriptionally engaged RNA polymerase and the production of mature mRNA transcripts (Kruesi et al. 2013). Mapping nascent RNA transcript reads relative to our active histone modification peaks revealed a dramatic alteration in the pattern of RNA polymerase elongation between embryos

and L1 larvae (Figure 3.5). Interestingly, Kruesi et al. noted that the transcription initiation sites of many genes were located at unexpectedly far distances upstream of the mature mRNA transcript; fortuitously, the mapping of DNA replication origins and active chromatin domains in the embryo provided us with landmarks to observe the genome-wide transition in transcription elongation between developmental stages (Figure 2.6; Figure 3.3).

Given the dramatic shift in the position of mRNA transcript production between early and late embryogenesis, we wondered whether the transition to the larval transcription pattern begins in embryos at the beginning of the morphogenesis transcription (Figure 2.6). The time series of nascent RNA sequencing libraries that we generated provided some novel insights into the regulation of RNA polymerase activity across embryogenesis (Figure 3.6).

From our observations, it appears that three principal mechanisms contribute to the shift in transcription following the end of mitotic cell divisions. Lineage-specific transcription factors that activate somatic cell genes 10-20 kb away from replication origins are expressed near the end of mitotic cell divisions, thus triggering the initiation of the morphogenesis program of gene expression (Hashimony et al. 2015). The binding of these transcription factors subsequently leads to the de novo remodeling of somatic gene promoters by active histone modifications, such as H3K27ac and H3K4me2, represented by the dynamic late peaks we have identified (Figure 3.2). The general remodeling of the transcriptional machinery to origin distal regions also appears to be accompanied by changes in RNA polymerase elongation

in *cis*, whereby the transcriptional activity shifts away from the constitutive chromatin domains associated with DNA replication origins (Figure 2.6; Figure 3.4). We propose that the cumulative action of these transcriptional remodeling mechanisms explains our initial observation that mRNA transcription relative to DNA replication origins undergoes a dramatic spatiotemporal shift that is initiated at the end of mitotic divisions (Figure 2.6).

A Developmental Context for Trans-Splicing in *C. elegans*

The tremendous concentration of transcriptional activity at constitutively transcribed promoters in germ cells and the early embryo may be designed to increase the efficiency of transcription required for the rapid processes of embryogenesis (Figure 3.4; Figure 3.6). However, the transition to a somatic program of gene expression for morphogenesis would require redistributing RNA polymerase and its associated transcription factors to somatic expressed genes located 10-15 kb away (Figure 2.6). While determining a mechanism to explain how this transition occurs is an active area of investigation, several models to explain our observations are proposed below.

One possible model invokes a specific developmental context for trans-splicing at *C. elegans* operons. Many of the genes which are constitutively transcribed from replication origin proximal promoters are operons and are characterized by clusters of 2-8 genes which can be transcribed as a single polycistronic pre-mRNA and processed as monocistronic mRNA transcripts through

the action of spliced leader trans-splicing (Blumenthal et al. 2015). The action of trans-splicing may stimulate the gradual elongation of RNA polymerase downstream of the 3' end of early expressed operons to the promoters at the 5' end of late expressed somatic genes (Figure 3.7).

This model would provide developmental context for the reported genetic and physical interactions between CstF-64, the *C. elegans* homolog of the mammalian RNA-binding subunit of mammalian 3' end formation factor CstF, and the SL2 snRNP (Kuersten et al. 1997; Evans et al. 2001). SL2 trans-splicing specificity for downstream genes in operons has been suggested to be coupled to the 3' end processing of upstream genes; weakening or strengthening of a polyadenylation signal in an upstream gene had a corresponding effect on the level of SL2 trans-splicing at the downstream gene (Kuersten et al. 1997; Liu et al. 2001). Another study described the purification of a complex containing CstF-64 and SL2 RNA; results showed that specific interaction between SL2 snRNPs and CstF-64 is dependent upon the SL2 RNA sequence, as mutation of a stem/loop completely abolishes the interaction and *in vivo* SL2 trans-splicing (Evans et al. 2001).

Our model proposing a function for SL trans-splicing in redistributing RNA polymerase to somatic cell differentiation genes in late embryos would explain the developmental phenotype reported for genes involved in pre-mRNA 3'-end formation and transcription termination (Cui et al. 2008). Mutations in the gene encoding the *C. elegans* homolog of SF2/ASF, a serine/arginine rich splicing factor, result in a lethal embryonic phenotype specifically during the morphogenesis phase (Longman et al.

2000). It is possible that the specific regulation of SL2 trans-splicing and transcription termination is required to shift a significant allotment of RNA polymerase to somatic gene promoters to activation during morphogenesis. We are currently investigating this model by examining how nascent RNA transcription patterns are altered by RNAi knockdown of additional splicing factors and 3' end formation factors.

Future Objectives: RNA Interference Pathways and Genome Regulation

In *C. elegans* embryos, the spatiotemporal gene expression programs that pattern the cell lineage for the entire animal require redundant regulatory mechanisms that stabilize transcription of developmentally appropriate genes. Thus far, we have discussed chromatin-mediated control of transcription initiation by histone acetylation and methylation, a developmental role for SL trans-splicing, and dynamic spatiotemporal patterns of RNA polymerase elongation. However, one of the prevailing gene regulatory adaptations in the natural history of *C. elegans* is undoubtedly the deployment of small RNA pathways to regulate gene expression. The discovery of RNA interference (RNAi) in *C. elegans* in 1998 revealed a ubiquitous epigenetic module that uses short RNAs, usually 20-30 nucleotides, to recognize and manipulate complementary nucleic acids (Fire et al. 1998, Obbard et al. 2008). The potency of the RNAi pathways derive from their ability to operate in trans in the nucleus and the cytoplasm, potentiating surveillance of RNA transcripts expressed from the entire genome (Grewal and Elgin 2007, Henderson and Jacobsen 2007).

The following review of RNAi in *C. elegans* is limited in scope to the 22G RNA pathways, which are the main effectors of small RNA gene regulation in *C. elegans* (Figure 3.8). While the complete developmental context for the small RNA pathways remains unknown, the core components of these pathways have been described, and serve a function in maintaining germ and soma-specific gene expression programs during the germline cycle and early embryogenesis. The introduction into RNA interference will be followed by an outline of proposed models and experiments to probe the role of *C. elegans* RNA interference pathways in regulating developmental gene expression in the embryo.

The Core RNAi Machinery

RNAi pathways are thought to have evolved as a defense against viral RNA genomes, but have been incorporated into the post-transcriptional regulation of endogenous gene expression in many animals, including *C. elegans* (Fire et al. 1998, Ketting et al. 2001). Mutations to any of the core genes involved in the *C. elegans* RNAi pathway results in severe developmental phenotypes including cell fate defects, embryonic lethality, or sterility (Grishok et al. 2001, Ketting et al. 2001, Smardon et al. 2000). During RNAi, members of the Dicer family of proteins process double stranded RNA (dsRNA) to initiate gene silencing (Fire et al. 1998; Carmell and Hannon, 2004). Dicer can process dsRNAs from exogenous or endogenous sources, generating small interfering RNAs (siRNAs) that guide sequence-specific silencing; biochemical studies revealed that Dicer and siRNAs are loaded into a large

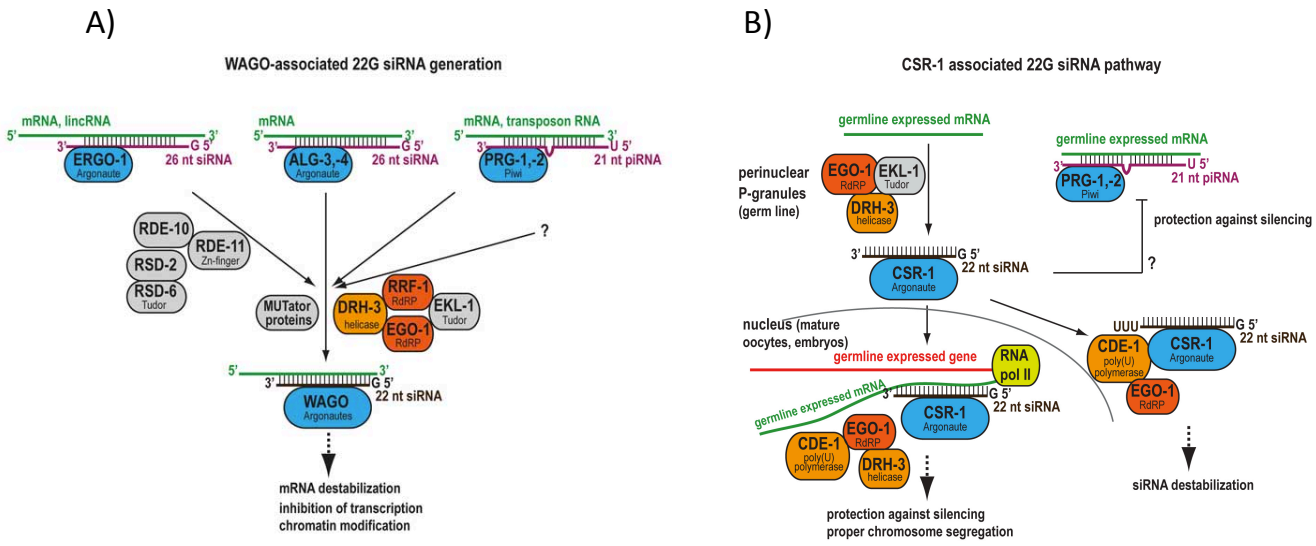

Figure 3.8. WAGO and CSR-1 associated 22G siRNA pathways. A) Effective RNAi in *C. elegans* relies on the amplification of primary RNAi triggers through the 22G RNA pathways. Three primary RNAi pathways, ERGO-1, ALG-3/4, and PRG-1/2, interact with their target and produce primary small RNAs, ERGO-1 and ALG-3/4 produce 26G RNA and PRG-1/2 produce 21U piRNAs– these small RNAs converge on the conserved module of the DRH-3 helicase, the EKL-1 Tudor domain protein, and a germline specific (EGO-1) or somatic specific (RRF-1) Argonaute to produce small 22G RNAs. These 22Gs interact with the WAGO clade of Argonautes to effect a majority of gene silencing in *C. elegans*. B) The CSR-1 22G pathway regulates expression of germline and early embryonic transcripts. The CSR-1 22G module relies on a germline specific RdRP named EGO-1 to produce 22G RNAs against germline transcirpts– CSR-1 bound 22G RNAs guide the Argonaute to nascent RNA transcripts produced in the nucleus, for the purpose of protecting the transcripts from the germline silencing 21U piRNA pathway, composed of PRG-1/2/. CSR-1 bound 22G RNAs are modified by the poly(U) polymerase CDE-1, which has an apparently destabilizing effect whose purpose remains unknown. From Billi et al. 2013.

complex called the RNA-induced silencing complex (RISC) (Liu et al. 2003, Pham et al. 2004, Tomari et al. 2004).

Endogenous RNAi Pathways

Studies have since revealed that the primary effectors of RNA silencing contained with RISC are members of the Argonaute (AGO) family (Liu et al. 2004, Meister et al. 2004). AGOs can be identified by the presence of a PIWI domain, which confers endonucleolytic "slicing" activity, PAZ, and MID domains, which are involved in 5' and 3' small RNA recognition (Liu et al. 2004). *C. elegans* encodes 24 AGO proteins, roughly half of which belong to a worm specific Argonaute (WAGO) clade (Fischer 2010). Once charged with a small RNA, AGO proteins are thought to mediate both the target sensing and effector steps of RNAi-related mechanisms (Song and Joshua-Tor, 2006). The antisense orientation of *in vivo* small RNAs associated with AGOs in *C. elegans* and the limited number of Dicer molecules suggested that the silencing must involve an amplification mechanism to increase the number of transcripts capable of screening by the Argonaute proteins (Tijsterman et al. 2002, Vasale et al. 2009).

The discovery of RNA dependent RNA polymerases (RdRPs) revealed that the majority of gene silencing by small RNA pathways depends on the direct synthesis of secondary siRNAs, called 22G RNAs, from the target mRNA transcript (Figure 3.8A) (Pak and Fire 2007). The production of 22G RNAs depends on a core module consisting of an RdRP (RRF-1 in somatic cells and CSR-1 in germ/embryonic

cells), a conserved Dicer-related helicase, DRH-3, and a Tudor domain protein EKL-1 (Aoki et al. 2007, Claycomb et al. 2009, Matranga et al. 2010). The production of 22G RNAs involves an AGO mediated interaction between the 3' end of the target mRNA the RdRP, followed by the de novo synthesis of 22G RNAs that are stripped off the template mRNA by DRH-3 and loaded onto its associated AGO for targeting (Pak and Fire 2007).

It is not known what directs RdRPs to initiate at target transcripts, however the targeting may be related to their production in Mutator foci (Phillips et al. 2012, Zhang et al. 2012, Yang et al. 2012) Mutator foci are perinuclear structures required for siRNA-mediated silencing in germline and somatic cells and are associated with a high concentration of RdRPs and WAGOs (Phillips et al. 2014, Conine et al. 2010, Batista et al. 2008). Based on a perinuclear association with the nuclear pore proteins of germ cell nuclei, it has been proposed that Mutator foci represent the primary sites for 22G RNA synthesis; in this model, WAGOs and RdRPs positioned at the nuclear pore are able to screen, amplify, or silence mRNA transcripts as they are exported to the cytoplasm (de Albuquerque et al. 2015).

CSR-1 22G RNA Pathway

Current estimates suggest that more than 50% of the endogenous transcriptome is targeted by 22G RNAs and many of these targets are exclusively expressed in germ cells or early embryos (Claycomb et al. 2009, Gerson-Gurwitz et al. 2016). A germline specific RdRP named EGO-1, produces a special class of 22G

RNAs bound by the only essential *C. elegans* Arogonaute, CSR-1 (Figure 3.8B) (Smardon et al. 2000, Claycomb et al. 2009). CSR-1 mutants display early embryonic lethality in the initial embryonic division cycles, as chromosomes fail to align at the metaphase plate and kinetochores do not orient to opposing spindle poles (Claycomb et al. 2009). Sequencing of CSR-1 bound small RNAs revealed that the pathway engages more than 80% of the germline transcriptome, including germline transcribed mRNA transcripts required for completion of meiosis and early embryogenesis (Wedeles et al. 2013).

The maternal loading of proteins and RNAs into oocytes has been shown to represent a conserved mechanism to regulate the transcriptional and translational activity during early embryogenesis (Baugh et al. 2003). Surprisingly, the CSR-1 pathway has been shown to broadly tune the level of maternal RNA transcripts in the early embryo through the selective cleavage of target gene transcripts—comparisons of WT and catalytically dead CSR-1 animals suggests that transcripts with greater numbers of homologous CSR-1 bound 22G RNAs are sliced to a greater extent than those with fewer 22G RNAs (Gerson-Gurwitz et al. 2016, Fassnacht et al. 2018). CSR-1 target genes include the majority of *C. elegans* operons that encode essential cellular functions in germ cells and somatic cells (Tu et. al 2014, Blumenthal et al. 2002).

Several outstanding questions regarding the CSR-1 pathway remain, including how CSR-1 22G RNA levels affect slicing activity and if CSR-1 acts in the post-

transcriptional regulation of gene expression at additional stages of *C. elegans* development.

WAGO 22G RNA Pathways Influence RNA Polymerase Dynamics

It is also possible that the transition in RNA polymerase elongation from the germ cell/embryonic gene expression state to the somatic differentiation state may involve small regulatory RNA pathways, specifically the WAGO 22G nuclear RNAi pathway (Billi et al. 2013). The WAGO 22G pathway executes RNA interference through the production of short 22 nucleotide small RNAs (22G RNAs) by RNA dependent RNA polymerases (RdRPs); these 22G RNAs guide the regulation of endogenous *C. elegans* transcripts by Argonaute proteins (Billi et al. 2013). *C. elegans* encodes 16 worm-specific Argonautes (WAGOs) that each are associated with specific 22G RNAs and endogenous transcript targets. The only essential Argonaute in *C. elegans* is CSR-1, which has a well characterized role in regulating the transcript levels in germ cells and early embryos, that was discussed in Chapter 1 (Yigit et al. 2006; Claycomb et al. 2009). Nuclear RNAi pathways are required for the trans-generational regulation of endogenous *C. elegans* transcripts and have previously been linked to the regulation of transcription elongation by RNA polymerase (Guang et al. 2010).

The core components of the WAGO 22G nuclear RNAi pathway include the SR-domain protein NRDE-2 and two nematode-specific nuclear proteins NRDE-1 and NRDE-4 (Burkhart et al. 2011; Buckley et al. 2012). This pathway has been

shown to inhibit RNA polymerase elongation by engaging with nascent pre-mRNA targets in a manner dependent upon a 22G RNA guide (Guang et al. 2010; Buckley et al. 2012). The exact mechanism by which RNA polymerase elongation is inhibited is currently unknown, but appears to involve the 22G guided association with pre-mRNA transcripts (Guang et al. 2010).

The fact that the *C. elegans* small regulatory RNA pathways have not been examined in a synchronous period of development, the extent to which these RNAi pathways regulate the endogenous protein-coding transcriptome remains largely unknown. In our model, we propose that RNA polymerase elongation is restricted during germ cell development and early embryogenesis by the action of the essential Argonaute CSR-1.

Model: CSR-1 Regulates Embryonic Transcriptome Dynamics

Current estimates suggest that more than 50% of the endogenous transcriptome is targeted by 22G RNAs and many of these targets are exclusively expressed in germ cells or early embryos (Claycomb et al. 2009, Gerson-Gurwitz et al. 2016). A germline specific RdRP named EGO-1, produces a special class of 22G RNAs bound by the only essential *C. elegans* Arogonaute, CSR-1 (Smardon et al. 2000, Claycomb et al. 2009). CSR-1 mutants display early embryonic lethality in the initial embryonic division cycles, as chromosomes fail to align at the metaphase plate and kinetochores do not orient to opposing spindle poles (Claycomb et al. 2009). Sequencing of CSR-1 bound small RNAs revealed that the pathway engages more

than 80% of the germline transcriptome, including germline transcribed mRNA transcripts required for completion of meiosis and early embryogenesis (Wedeles et al. 2013).

The maternal loading of proteins and RNAs into oocytes has been shown to represent a conserved mechanism to regulate the transcriptional and translational activity during early embryogenesis (Baugh et al. 2003). The CSR-1 pathway has been shown to broadly tune the level of maternal RNA transcripts in the early embryo through the selective cleavage of target gene transcripts—comparisons of WT and catalytically dead CSR-1 animals suggests that transcripts with greater numbers of homologous CSR-1 bound 22G RNAs are sliced to a greater extent than those with fewer 22G RNAs (Gerson-Gurwitz et al. 2016, Fassnacht et al. 2018). This mechanism of gene silencing activity may allow CSR-1 to maintain the homeostasis of the early embryonic transcriptome. This may be required, as CSR-1 target genes include the majority of *C. elegans* operons that encode essential cellular functions such as growth and proliferation in germ cells and somatic cells (Tu et. al 2014, Blumenthal et al. 2002).

Given the association of the constitutive peaks with maternal RNA transcripts detected in the early embryo, we decided to compare the distribution of CSR-1 target genes relative to constitutive peak midpoints. Using the CSR-1 target gene lists from Claycomb et al. 2009 and 4-cell embryo transcript data from Boeck et al. 2016, we were able to plot the RNA count of CSR-1 targeted transcripts relative to the midpoints of constitutive, dynamic early, or dynamic late peaks (Figure 3.3). This

analysis revealed a nearly 2-fold enrichment of CSR-1 targeted transcripts proximal to constitutive peaks which we find to be associated with DNA replication origins (Figure 3.9). The Claycomb lab has reported that CSR-1 preferentially associates with nascent RNA transcripts at chromosomal regions marked by active histone modifications found at gene promoters, such as H3K4 methylation and H3K27 acetylation (Wedeles et al. 2013). Indeed, we find that CSR-1 target genes are specifically associated with constitutively accessible chromatin regions (marked with active histone modifications) that specify DNA replication origins during embryogenesis. While this association is striking, it currently remains unknown whether the association between CSR-1 target genes and constitutively accessible chromatin domains reflects an essential genomic arrangement or merely is coincidental. Recent analysis has suggested that CSR-1 target genes are conserved between *C. elegans* and *C. briggsae*, a related nematode species; it would be of great value to analyze whether CSR-1 target genes are associated with DNA replication origins and accessible chromatin regions in *C. briggsae* (Tu et al. 2014). Such an association may point to a coevolutionary relationship between germline expressed CSR-1 target genes and DNA replication origins.

While there has been considerable attention given to CSR-1 because of its essential role in regulating germ cell development and the deposition of maternal RNA in oocytes, there have been limited investigations as to how CSR-1 functions in

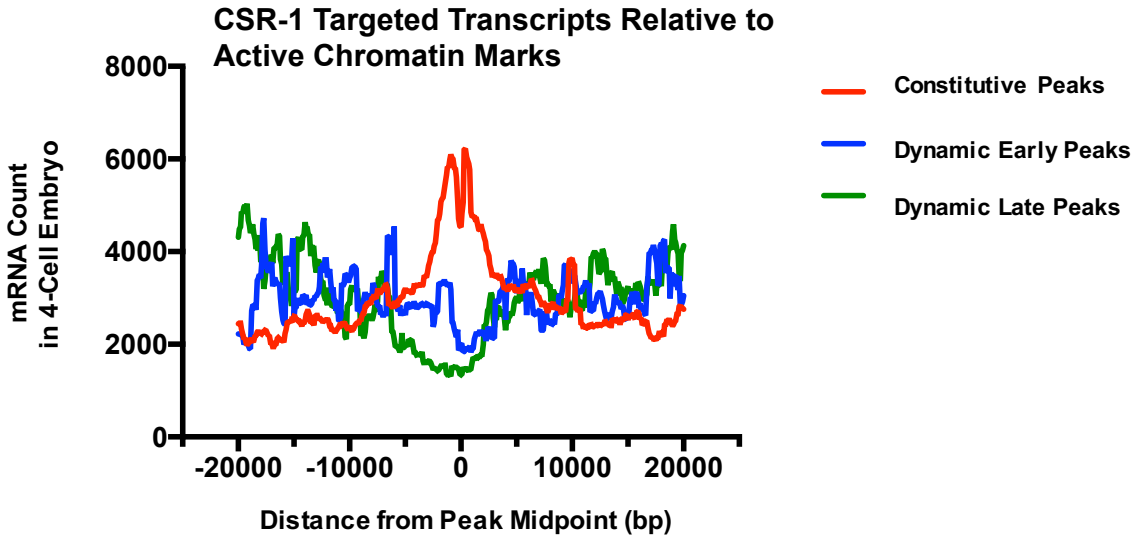

Figure 3.9. CSR-1 target genes are enriched at domains of embryonic transcription remodeling. 4,191 CSR-1 target genes were identified by sequencing CSR- associated 22G RNAs (data from Claycomb et al. 2009); the transcript counts of CSR-1 target genes in the 4-cell *C. elegans* embryo (data from Boeck et al. 2016) are mapped in 40 kb genomic ranges around the midpoints of dynamic early (H3K4me2 and H3K27ac; n=3541), dynamic late (H3K4me2 and H3K27ac; n=3381), and constitutive (overlapping H3K27ac and H3K4me2; n=4180) active histone modification peaks.

the developing embryo. Guided by our maps of nascent RNA transcription and chromatin modifications, we observe that:

1) CSR-1 target gene promoters retain H3K27ac and H3K4me2 marks from the germline through the entirety of embryo development (Figure 3.9)
2) RNA polymerase transcription from CSR-1 target gene promoters can be detected as early as the 4-cell stage by nascent RNA sequencing, suggesting that the first zygotic transcripts interact with CSR-1
3) CSR-1 target transcripts appear to be transcribed but not expressed in the terminally differentiated cells of the late embryo

These findings provide an opportunity to examine how CSR-1 activity affects the embryonic transcriptome, particularly whether the dynamic remodeling of RNA polymerase elongation from early to late embryogenesis is a result of CSR-1 22G pathway regulation. While CSR-1 has not been studied in the context of embryogenesis, the expansive literature pertaining to endogenous RNAi pathways provides some likely models which fit with our data. In light of this new data, we would specifically like to address several new questions in our impending studies:

1) Is CSR-1 is responsible for maintaining the germline transcription patterns that persists in the embryo (Figure 3.4C) from fertilization to ~300 minutes, after which gene transcription is remodeled (Figure 2.5; Figure 3.4B)?

2) Does CSR-1 play a role in regulating RNA polymerase dynamics, as has been proposed for nuclear RNAi pathways (Guang et al. 2009)?

3) Does the titration of CSR-1 protein during embryogenesis result in the gradual shift from replication origin-proximal transcription to origin-distal transcription in the late embryo?

4) How are growth genes transcribed (Figure 3.6; green) but not expressed in late embryos and do RNAi pathways have any role in the silencing of these transcripts?

CSR-1 target genes represent ~80% of the germline transcriptome, as well as the majority of maternal RNA transcripts deposited into oocytes (Fassnach et al. 2018). Early studies of CSR-1 suggested that instead of silencing its target transcripts, CSR-1 appears to play a role in licensing their expression (Claycomb et al. 2009); paradoxically though, a majority of the catalytic Argonaute slicing activity purified from *C. elegans* extracts comes from CSR-1 (Aoki et al. 2007). To determine a role for the catalytic slicing function of CSR-1 in early embryos, the Desai lab has

expressed a “slicing-deficient” copy of CSR-1 to identify effects on embryo viability and maternal RNA transcripts (Gerson-Gurwitz et al. 2016). Their results indicate that CSR-1 does not completely silence target genes which are complementary to small 22G RNAs—instead, the mutant studies reveal that CSR-1 broadly tunes the expression of its target genes through selective slicing activity (Gerson-Gurwitz et al. 2016). CSR-1 slicing is thought to be modulated by the density of homologous small RNAs, whose generation is directly related to the target transcript abundance—highly abundant transcripts trigger increased production of small RNAs and an increase in CSR-1 slicing, while lowly abundant mRNAs correspond to low small RNA levels and less CSR-1 slicing (Gerson-Gurwitz et al. 2016). While this model appears to clarify the role for CSR-1 in regulating maternal RNA transcripts, it would be difficult to elucidate the mechanism of CSR-1 in the embryo without a synchronous time course or evaluation of nascent pre-mRNA.

To test for a possible role for CSR-1 in regulating embryonic transcription, we will use temperature-sensitive mutants which abolish the production of the small RNAs required for CSR-1 association with target mRNAs. This approach avoids the embryonic lethality associated with CSR-1 RNAi (Claycomb et al. 2009) and allows us to evaluate the phenotypes of differently staged embryo populations. We plan on using a temperature sensitive allele for the Dicer-related helicase *drh-3*, which encodes a DEAH/D-box helicase required for germline RNAi and the production of 22G small RNAs (Tabara et al. 2002). By shifting *drh-3* gravid adults to the non-

permissive temperature prior to extracting the embryos, we will be able to isolate viable, fertilized embryos for mature mRNA and nascent pre-mRNA sequencing analysis. The prediction is that if CSR-1 association with its target transcripts is required to stabilize early embryo transcription patterns, the *drh-3* mutant will show aberrant gene expression by nascent RNA sequencing and mRNA sequencing. If CSR-1 is acting to regulate RNA polymerase elongation, the Gro-seq analysis of early embryos should reveal the shifted pattern—if CSR-1 or another WAGO relies on 22G RNAs homologous to early expressed growth genes, then we would predict the expression of growth genes in late embryos when these transcripts appear to be silenced (Figure 2.5; left panel, 500-850 minutes). This initial experiment using the temperature sensitive *drh-3* allele might also provide a clue as to how CSR-1 may regulate embryonic transcription patterns. One particularly appealing possibility is that CSR-1 Argonaute and its 22G RNAs directly influence RNA polymerase elongation through the nuclear RNAi pathway.

In such a model, the spatiotemporal restriction of RNA polymerase elongation by CSR-1 may function as a mechanism to concentrate RNA polymerase and co-transcriptional RNA processing factors at genes which are rapidly transcribed during maternal RNA synthesis and early embryogenesis. We suggest that due to the exclusive expression of CSR-1 in germ cells, the concentration of nuclear CSR-1 might become titrated away during early embryonic DNA replication and cell divisions, such that transcription elongation becomes more pervasive as

embryogenesis progresses and the effect of CSR-1 is attenuated. Given the strict specificity governing the association between 22G RNAs and WAGOs, the regulation of transcription elongation can be restricted specifically to the genes targeted by CSR-1 bound 22G RNAs (Gerson-Gurwitz et al. 2016; Claycomb et al. 2009). We also plan on specifically testing whether the nuclear RNAi pathway components, NRDE-2/4/1, regulate embryonic transcription by generating nascent RNA sequencing libraries from a temperature sensitive allele of *nrde-2* (Burkhart et al. 2011). While the specific mechanism that inhibits RNA polymerase elongation is not known, the *nrde-2* mutant was previously shown to exhibit abberrant "read-through" transcription of the *lin-15A/lin-15B* operon when evaluated by Gro-seq (Buckley et al. 2012). Our expectation is that a Gro-seq analysis of the temperature sensitive *nrde-2* and the *drh-3* mutant embryos should reveal whether RNA polymerase elongation in during embryogenesis relies on the endogenous RNAi pathway components to regulate spatiotemporal patterns of gene expression.

Furthermore, we propose that if the pattern of RNA polymerase elongation depends on CSR-1 and 22G RNAs association with nascent transcripts in the nucleus, that would suggest a reliance of DNA replication to gradually transition to the program of somatic lineage specific gene expression that has been reported previously (Edgar & McGhee, 1988). This model would indicate a novel spatiotemporal role for the essential CSR-1 22G RNA pathway in regulating the transition between germ and somatic gene regulatory controls.

Chapter 4

Conclusion

DNA Replication and Genome Organization

The propagation of genomic information hinges upon its accurate duplication during every cellular division—early animal development is largely devoted to the process of regulating DNA replication to expand the early cell population (Murray and Kirschner, 1989, Edgar et al. 1994). The results published in this work highlight a novel role for DNA replication origins in organizing the *Caenorhabditis elegans* genome. We propose that the evolution of a DNA replication landscape to support the hermaphrodite life cycle of the worm incurred a developmental transcriptome that is optimized for rapid embryogenesis and a high reproductive capacity.

Evidence from species throughout the phylogenetic tree suggest that DNA replication provides the first imprint of the genome structure that patterns the cascade of differential regulatory events during early embryo development (Pourkarimi et al. 2016, Ke et al. 2017, Hug et al. 2017). Chromosome conformation maps obtained from mouse gametes, zygotes, and 8-cell embryos demonstrated that the establishment of TAD structures requires DNA replication, as G1 arrested 2-cell embryos lack any discernible TAD boundaries or clustered contacts (Ke et al. 2017). In *Drosophila* embryos, it was observed that the sharpening of TAD boundaries during early embryogenesis is dependent upon DNA replication in order to establish

the broad activation of zygotic transcription at the mid-blastula transition (Hug et al. 2017). Altogether, these findings support the conclusion that the formation of regulated chromosomal domains of gene expression is related to the gradual acquisition of defined replication domains preceding zygotic transcriptional activity (Rivera-Mulia et al. 2015).

Yet, how do DNA replication origins become defined over the course of genome evolution? It has been widely noted that efficient DNA replication origins from yeast to human are associated with highly transcribed genes encoding essential cell viability functions, including transcription, RNA metabolism, and chromosome segregation (Gilbert et al. 2010). The non-random positioning of essential genes proximal to replication origins (including genes encoding DNA replication factors themelves) has been proposed to be related to the lower mutation rate in early replicating regions and the increased rate of mutation and evolutionary divergence in later replicating regions (Chuang & Li, 2004). According to this model, genes in mutationally hot regions (i.e. late replicating) are biased towards differentiation and extracellular communication, while genes in cold regions (i.e. early replicating) are biased towards essential cellular processes, like gene regulation, RNA processing, and protein modification (Stamatoyannopoulous et al. 2009; Chen et al. 2010). This gene organization may have been selected to minimize the mutational load on genes that need to be conserved and allow fast evolution for genes that must frequently adapt.

To test the idea that DNA replication influences mutation rate across the genome, we decided to measure the distribution of naturally occurring sequence variants (allele frequency <1% in population) among globally distributed *C. elegans* isolates relative to our mapped DNA replication origin positions. The choice to map only low frequency sequence variants serves to exclude mutations which have been selected for within highly inbred populations, as these alleles may have provided fitness advantage and become fixed within the population. The sequence variations were compiled from whole genome sequencing of 29 *C. elegans* isolates (Cook et al. 2016); using VCF tools, the alleles represented in less than 1% of the populations were filtered and the median score was reported in 100 bp genomic ranges, generating a genome-wide distribution of natural sequence variation. When we plotted the median allele frequency at varying distances from efficient (>0.5) DNA replication origins, we observed a linear relationship between mutation rate and distance from DNA replication origins (Figure 4.1). This analysis suggests that DNA sequences near replication origins may be less inclined to mutate, which creates a "safe harbor" for essential genes near replication origins. Across eukaryotic species, housekeeping genes, which are broadly expressed across cell types and tissues, have been found to be associated with DNA replication origins (Dixon et al. 2016, Pourkarimi et al. 2016, Petryk et al. 2016). However, it remains unclear which genomic feature, evolutionarily speaking, is causal of the association; disentangling essential gene transcription from DNA replication is challenging as disruption of one might impact the activity of the other. Determining the relationship between DNA

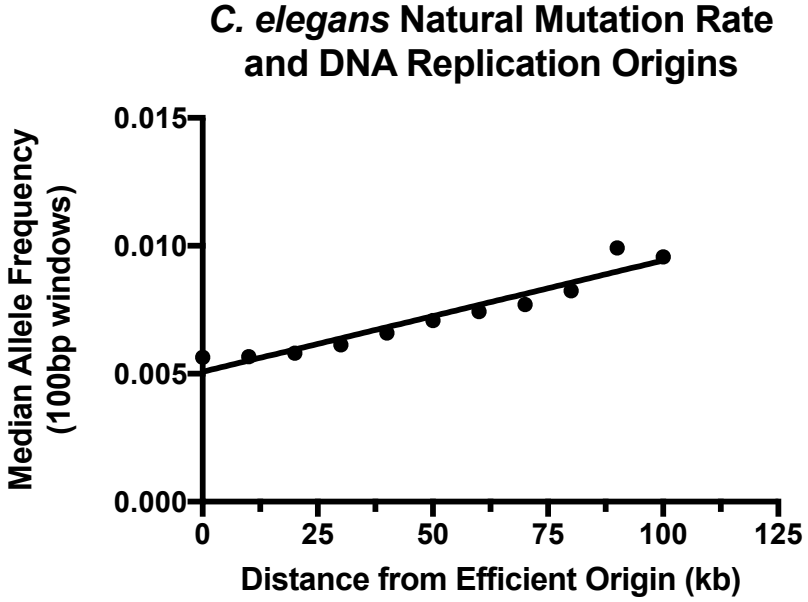

Figure 4.1. Natural mutation rate across *C. elegans* isolates is influenced by proximity to DNA replication origins. Using VCF tools, allele frequencies at each nucleotide were calculated for the 29 sequenced genomes and filtered for "neutral sequence variants" represented at less than 1% among the isolates. Among the natural variants, the median allele frequency was calculated in 100 bp genomic windows (Y-axis) and plotted at varying distances from efficient origins (X-axis). Data curated by Cook et al. 2016 and retrieved from CeNDR (*C. elegans* Natural Diversity Resource).

replication landscapes and gene evolution remains an intriguing avenue for investigation, especially in the context of understanding the driving forces behind metazoan genome evolution.

While we find that DNA replication timing has an influence on mutation rate across the genome, only germline mutations are passed on to offspring. Given this fact, it is reasonable to propose that germline-specified DNA replication timing might influence the distribution of gene sequences across the genome. The continuity of the germ line may depend more on the sequence conservation of essential viability genes, involved in gene regulation, RNA processing, and protein homeostasis, than it does on differentiation genes for specialized somatic cell function, thus skewing the gene distribution. Most of the essential genes in *C. elegans* are germline expressed, suggesting that DNA replication patterns in the germ cells may be critical for shaping *C. elegans* genome organization (Reinke et al. 2000, Kamath et al. 2003). The consistent pattern of DNA replication we observe at all stages of *C. elegans* embryogenesis suggests that origins may be stably passaged from the adult germline to the embryo, thus protecting the integrity of essential germline genes and the propagation of the species.

Germline Transcription, Operons, and Rapid Embryogenesis

Nematodes represent one of the few animal genomes that organize genes into operons, enabling the transcription of 2-8 genes from a single promoter (Spieth et al.

1993). Early identification of cDNA clones containing SL2 RNA leader sequences revealed that most genes which were co-transcribed in operons encode conserved functions interacting within the same molecular pathway—representative gene groups within operons are involved in transcription, splicing, RNA degradation, apoptosis, ribosome biogenesis, histone acetylation, proteasomal degradation, and cyclophilins (Blumenthal and Gleason 2003).

Not surprisingly, these operons are consolidated around DNA replication origins and are among the earliest transcribed genes in the embryo (Pourkarimi et al. 2016). We hypothesize that the ability to process polycistronic RNA transcripts from operons through SL trans-splicing greatly increased the efficiency with which diverse housekeeping genes could be expressed during early embryonic divisions.

Operons proximal to replication origins experience more potent transcriptional burst due to their early replication and increased copy number for the duration of the cell cycle. The greater transcriptional efficiency attributed to operon structure and SL trans-splicing may also explain the absence of a cell cycle pause characteristic of the mid-blastula transition during *Drosophila* and *Xenopus* embryogenesis (Lee et al. 2014). The coupling the transcriptional program for cellular proliferation near origins is particularly well suited for the “boom” and “bust” developmental life cycles of wild *C. elegans* populations, enabling populations to toggle between rapid extremes of reproduction and stress tolerance in between periods of feeding (Felix and Braendle, 2010).

The coupling of operon transcription to DNA replication origins provides a logical temporal switch to increase gene expression to support growth and proliferation; this might be particularly important during *C. elegans* germline development (Kimble and White, 1981). In wild populations, the ability to regenerate the germline quickly after periods of long starvation would be advantageous to boost reproduction when a new food source is found (Blumenthal and Gleason, 2003). Germline proliferation involves rapid division of mitotic germ cells during the L1-L2 stages coupled with the transcription of genes required for proper germ cell development (Kimble and White, 1981, Crittenden et al. 1994, Tzur et al. 2018). It is possible that the ability to regenerate the germline to resume reproduction after long periods of stress was aided by the coupling of proliferation genes (encoded by germline-expressed operons) to DNA replication origins (Reinke et al. 2009, Seidel and Kimble, 2011).

It is plausible that the ~4000 constitutive peaks we find to be correlated with DNA replication origin efficiency, germline transcription, and conserved operons represent master control regions that are heritably maintained through germline cycles across *C. elegans* generations. Minimizing the regulatory capacity of the genome to a limited number of chromatin domains may have been driven by a natural history marked by extreme “feast” and “famine” cycles. In an organism that oscillates between long periods of starvation and minimal reproductive investment and periods of rapid growth and reproduction, the ability to rapidly toggle between these developmental programs may have been critical for *C. elegans* evolution.

Chapter 5

Methods

C. elegans Embryo Culture: Maintenance on Plates

All C. elegans strains were maintained at 20°C on NGM plates with OP50 E. coli strain, as previously described (Brenner, 1974). Briefly, for the large scale culture of worms, 10-15 NGM plates were seeded with 1 mL of 40X concentrated OP50 *E. coli*. Strains used in this study include the wild-type strain, corresponding to Bristol N2 (RRID:CGC_N2) and CG21 egl-30(tg26) I; him-5(e1490) V (RRID:CGC_CG21).

Worms were seeded on NGM plates at a density of 25,000 synchronized L1 larvae per plate. L1 larvae were obtained by washing mixed populations of worms from NGM plates with M9 buffer (1 mM MgCl2) and filtering with Millipore 11 uM Nylon filters.

Okazaki Fragment Sequencing

C. elegans Liquid Culture and RNAi

We used lig-1 RNAi generated by the Ahringer laboratory and used a modified version of RNAi feeding protocol (Fraser et al., 2000). lig-1 and control (empty vector) RNAi-expressing bacteria were grown overnight at 37°C in LB medium supplemented with 50 µg/ml Ampicillin and 10 µg/ml tetracycline. Bacterial cultures were transferred to fresh LB medium supplemented with 50 µg/ml Ampicillin and were grown at 37°C

until reaching OD 1.0. RNAi expression was induced with 1 mM IPTG for 1 hr at room temperature.

To attenuate lig-1 associated genome instability and prevent sterility associated with complete lig-1 depletion, lig-1 RNAi expressing bacteria was diluted 1:10 with control (empty vector) bacteria to a final volume of 500 ml. Diluted bacteria were harvested by centrifugation and added to C. elegans S-Basal liquid media containing 150,000–200,000 well synchronized early L3 staged N2 worms (see below). N2 worms were grown in S-Basal liquid culture containing diluted lig-1 RNAi at 20°C until adulthood. Pre Gastrula embryos were harvested from early L3 staged egl-30 mutants that were grown on 1 mM IPTG plates as described previously (Fraser et al., 2000). Gravid adults were harvested and their embryos were collected by bleaching. Additional lig-1 dilution ratios (Figure 1a) were generated as above.

To collect synchronous L1 populations, 10,000 worms were grown on NGM plates. Their progeny were washed with M9 buffer and passed through 11 µM nylon net filters (Millipore Ltd. NY1104700). Approximately 150,000 synchronized L1 stage worms were grown in liquid media supplemented with OP50 bacteria until reaching L3 stage. Synchronized L3 staged worms were collected by centrifugation, followed with three times washing with M9 buffer containing 50 µg/ml Ampicillin. For progeny counting L3 staged worms where grown on 1 mM IPTG plates.

Genomic DNA Purification and Okazaki Fragment Labeling

Okazaki fragments were purified as previously described with slight modifications (Smith and Whitehouse, 2012). Samples were resuspended in 480 µl Lysis buffer (50 mM Tris-HCL, pH 8.0, 50 mM EDTA, 100 mM NaCl, 1.5% Sarkosyl, 1% SDS) and incubated with 200 µg proteinase K at 42 degree overnight. Digested proteins and peptides were precipitated by addition of 200 µl 5 M KOAc and centrifugation at 16,000 g for 30 min at 4°C. Genomic DNA was precipitated by adding 500 µl isopropanol and spinning at 16,000 g at 4°C for 10 min. Genomic DNA pellets were washed with 70% ethanol and then resuspended in 300 µl STE (10 mM Tris-HCL, PH8.0, 1 mM EDTA, 100 mM NaCl). Residual RNA was digested by addition of 5 U RiboShredder RNase Blend (Epicentre) at 37°C overnight. Genomic DNA was precipitated with 30 µl 3 M NaOAc and 1.7 ml ethanol pelleted at 10,000 g for 10 min at 4°C. DNA pellets were washed with 70% ethanol and resuspended in 30 µl TE (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) and stored over night at 4°C to allow complete resuspension of genomic DNA. Samples were stored at −80°C. Radiolabeling of Okazaki fragments and denaturing gel electrophoresis were followed as previously described (Smith and Whitehouse, 2012). Input genomic DNA for labeling was 2–3 µg.

Okazaki Fragment Purification and Sequencing Library Generation

Okazaki fragments were purified from genomic DNA of Ligase I depleted C. elegansembryos, by ion exchange chromatography, similar to the procedure

previously described (Smith and Whitehouse, 2012). C. elegans Okazaki fragments were enriched in the 750–850 mM NaCl fractions. These fractions were pooled and DNA precipitated for sequencing library preparation.

Three hundred nanograms purified Okazaki fragments were used to generate sequencing libraries similar to the previously optimized protocol (Smith and Whitehouse, 2012). Following fragment ligation the total reaction was loaded on a 2% agarose gel; fragments corresponding to ~200–700 bp were purified from the gel using QiAquick gel extraction kit (Qiagen). Purified ligated fragments were amplified by (PCR 16) cycles using custom Illumina or Ion Torrent sequencing oligos.

Genomics Protocols

Sequencing was performed using Ion Torrent (Proton) or Illumina (HiSeq) platforms. At least one biological replicate was performed for each DNA replication map; Ion Torrent (Proton) generally had a higher background and lower reproducibility than Illumina so direct comparisons between different embryonic stages were made using only the Illumina data (Figure 3). Sequencing reads were mapped to the WS220 genome using Bowtie2 with –local function (RRID:SCR_005476). Reads with $q < 30$ were removed using Samtools (RRID:SCR_002105). Remaining reads were binned in 100 bp intervals using Bedtools (RRID:SCR_006646), maintaining strand identity. Data were partially smoothed by calculating the median with a sliding window of 1.5 kb (Royce et al.,

2007). Data were normalized (such that the sum of Watson and Crick reads = 1) and positions of origins were mapped using custom program described earlier, using a 12 kb window (McGuffee et al., 2013). Because replication initiates in a broad zone at most origins, we found that the method of McGuffee significantly underestimated the true origin efficiency; therefore, origin efficiencies are defined by the maximum value of normalized reads within 20 kb of the origin. The distance between normalized maxima on the Watson and Crick strands defined the size of transition zone at origins. Histone modification data used in Figure 2 were downloaded from ModEncode consortium (http://www.modencode.org/). Regions of ChIP enrichment were calculated using Macs2 (Zhang et al., 2008) with default parameters. Regions of enrichment were defined as peaks with >5 fold enrichment over background. Associations of ChIP peaks with replication origins was calculated using intervalstats (Chikina and Troyanskaya, 2012); replication origins were defined as a 5 kb region centered on the origin midpoint. Transcriptomics data from the whole embryo time course from (Hashimshony et al., 2015) were used in this study. Data for each gene were individually normalized such that the sum of all 50 time points = 1. Data were then clustered (unsupervised) and genes with highest transcription within the first five time points were defined as ‘maternal’ all other expressed genes are defined as ‘zygotic’.

Synchronized Embryo Timecourse for ChIP-sequencing

Embryo Collection

Approximately 10-15 NGM plates of gravid egl-30 (tg26) hermaphrodite adult worms were washed from plates with M9 buffer into 50 mL Falcon tubes and allowed to settle by gravity. The supernatant was removed and washes were repeated a total of 5 times until no embryos are visible in the supernatant. Embryos were bleached by vortexing until adult tissues were no longer visible; the embryos were washed 3X in 50 mL Falcon tubes with M9 buffer and resuspended in 3 mL of egg buffer (Edgar et al. 1988).

In vitro Culture of Synchronized Embryos for ChIP Seq

For each embryonic time point to be taken, approximately 60,000-100,000 embryos were aliquoted in 1 mL of egg buffer to 2 mL Eppendorf tubes. The synchronized embryo time course was carried out in a 20 degree incubator on a rotating stand; at the time point of embryo harvest, tubes were centrifuged for 1 minute at 10,000 g, washed 2X with 500 uL of M9 buffer, and then resuspended in 500 uL of M9 + 2% formaldehyde (37%) for cross-linking for 30 minutes.

Formaldehyde was quenched by adding glycine to a final concentration of 125 mM and incubated for 5 min. Embryos were washed 2X in 1 mL PBS + protease inhibitors and stored as dry pellets at -80 deg C until extract preparation.

Embryonic ChIP Extract Preparation and Immunoprecipitation

Embryos were resuspended in ~100 µl of embryos (volume estimated when embryos are collected) in 500 µl FA buffer + 0.1% sarkosyl* + protease/phosphatase inhibitors, dounced on ice for 30 strokes using a type B homogenizer pestle, and sonicated with a Bioruptor on settings: 4 deg C, 15 min on High, 30 sec on 30 sec off.

Samples were then transferred to microfuge tubes and spun at top speed for 15 min at 4 deg. Celsius. The supernatant was aliquoted and stored at -80 deg C until chromatin immunoprecipitation. 2 mg of embryonic extract was used per IP with 2 ug of antibody added to each reaction. The antibody was conjugated to 40 uL of Protein G beads, racked on a magnetic stand, and washed 3X with FA buffer. 500 uL aliquots of embryonic ChIP extract was added to the antibody conjugated beads and incubated overnight at 4 deg. C.

Immunocomplexes bound to beads were washed with FA buffer 2X for 5 min each, 1X with FA + 1 M salt for 5 min, 1X with FA + 500 mM salt for 10 min, 1X with TEL buffer for 10 min, and 2X with TE for 5 min each before purifying the beads and proceeding to sequencing library preparation.

Sequencing Library Preparation

Libraries were prepared on beads using the NEB Next ChIP-seq Library prep kit (NEB # E6200S) according to manufacturers instructions. Final PCR amplification was performed with KAPA 2X Hot Start Polymerase Master mix following the program of:

- 37 deg: 15 minutes (USER Digestion)
- 98 deg: 1 minute
- 98 deg: 15 sec
- 65 deg: 15 sec
- 72 deg: 1 min
- 72 deg: 5 min (final extension)
- Total of 16-18 cycles

Libraries were sequenced on Illumina Hi-Seq platforms to a depth of 10-20 million reads per sample.

Synchronized Embryo Timecourse for Gro-Seq

Embryo Extraction

For each embryonic time point to be taken, approximately 60,000-100,000 embryos were aliquoted in 1 mL of egg buffer to 2 mL Eppendorf tubes. The synchronized embryo time course was carried out in a 20 degree incubator on a rotating stand; at the time point of embryo harvest, tubes were centrifuged for 1 minute at 10,000 g, washed 2X with 500 uL of M9 buffer**.**

After washing twice with M9 buffer, animals were washed with cold nuclear isolation buffer (250 mM sucrose, 10 mM Tris-HCl (pH 7.9), 10 mM MgCl2, 1 mM EGTA, 0.25% NP-40, 1 mM DTT, protease inhibitors, 4 U/ml SUPERaseIn [AM2696;

Ambion, Grand Island, NY]). Animals were resuspended in nuclear isolation buffer (embryos and starved L1 in 3 vol, L3 in 1 vol), and dripped into liquid nitrogen to freeze. Starved L1 and L3 samples were ground under liquid nitrogen by mortar and pestle. Larval samples, post-grinding, and embryo samples were dounced with a Kontes 2 ml glass dounce to release nuclei. Douncing and collection of nuclei was performed for up to six rounds as follows: dounce with 10X pestle A, 10X pestle B, 5 min centrifugation at 100×g, removal of nuclei-containing supernatant, and addition of an equal volume of nuclear isolation buffer to the pellet. Nuclear isolation was monitored each round to determine effectiveness and when it was complete. The pooled supernatant was centrifuged for 5 min at 1000×g to pellet nuclei. The nuclear pellet was washed with nuclear freezing buffer (40% glycerol, 50 mM Tris-HCl (pH 8.3), 0.1 mM EDTA, 5 mM MgCl2, 1 mM DTT, protease inhibitors, 4 U/ml SUPERaseIn). Approximately 1 × 108 nuclei were resuspended in 100 µl nuclear freezing buffer and stored at −80°C until GRO-seq reactions were performed.

Preparation of Gro-Seq Libraries

NRO Reaction

Nuclei (100 µl) were mixed with an equal volume of reaction buffer (10 mM Tris-HCl (pH 8.0), 5 mM MgCl2, 1 mM DTT, 300 mM KCl, 20 U of SUPERaseIn, 1% sarkosyl, 500 µM each of ATP, GTP, and Br-UTP, 2 µM CTP, and 0.33 µM α-32P-CTP [3000 Ci/mmol]). The reaction was allowed to proceed for 5 min at 30°C. The reaction was stopped by the addition of 2 ml (10× volume) of TRIzol (Invitrogen). The

phases were separated by the addition of 400 µl of chloroform as per the manufacturer's instructions. An additional acid-phenol and then chloroform extraction were carried out, followed by precipitation with 2.5 vol of ethanol. The pellet was washed in 75% ethanol before resuspending in 20 µl of DEPC-treated water. Base hydrolysis was performed on ice by the addition of 5 µl 1 M NaOH and incubated on ice for 30 min. The reaction was neutralized by the addition of 25 µl 1 M Tris-HCl (pH 6.8). The reaction was then run through a p-30 RNAse-free spin column (BioR, Hercules, CA) according to the manufacturer's instructions. The column flowthrough was brought to 100 µl with DEPC water and EDTA was added to a final concentration of 1 mM.

Bead pre-wash

All buffers used in bead enrichment steps were kept on ice and were supplemented with 4 U/ml of SUPERaseIn. Anti-deoxyBrU beads (#sc-32323-ac; Santa Cruz Biotech, Santa Cruz, CA) were first washed three times with a pre-wash buffer: 0.25× SSPE, 500 mM NaCl, 1 mM EDTA, 0.05% Tween for 5 min; washed twice in binding buffer: 0.25× SSPE, 37.5 mM NaCl, 1 mM EDTA, 0.05% Tween for 5 min; blocked in bead blocking buffer: 0.25× SSPE, 1 mM EDTA, 0.05% Tween, 0.1% PVP, and 1 mg/ml ultrapure BSA (AM2618; Ambion) for 1 hr; followed by one wash in binding buffer for 5 min. The ratio of beads to volume did not exceed 1:8 for any wash or blocking step. The beads were resuspended in a 25% slurry (original concentration).

Bead enrichment

NRO RNA was heat denatured at 70°C for 3 min and placed on ice for 2 min. Then, 350 µl of binding buffer and 50 µl of bead slurry were added to the RNA, and the samples were incubated for 30 min on a rotating stand (8 rpm). The beads were washed once in binding buffer; once in low salt wash buffer: 0.2× SSPE, 1 mM EDTA, 0.05% Tween; once in high salt wash buffer: 0.25% SSPE, 137.5 mM NaCl, 1 mM EDTA, 0.05% Tween; and twice in TET: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.05% Tween. The NRO RNA was eluted three times (2× 125 µl, 1× 250 µl) with elution buffer: 20 mM DTT, 150 mM NaCl, 5 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 0.1% SDS. The NRO RNA was then isolated by a standard extraction-precipitation method: one acid-phenol extraction, one chloroform extraction, addition of NaCl to 300 mM and 1 µl of glycoblue (AM9515; Ambion) to the aqueous phase, precipitation with 2.5 vol of cold ethanol, and a wash of the resulting pellet with 75% ethanol. The pellet was resuspended in DEPC water at volumes appropriate for the subsequent step.

Sequencing Library Preparation

Sequencing libraries were prepared using the SMARTer Stranded RNA-seq Kit according to manufacturers instructions. PCR amplification was performed using KAPA Hi-Fi Polymerase 2XMaster Mix with the following PCR Program:

- 98 deg: 1 minute
- 98 deg: 15 sec
- 65 deg: 15 sec
- 72 deg: 1 min
- 72 deg: 5 min (final extension)
- Total of 10-12 cycles

Reads were mapped to the WS235 genome and trimmed on the 5’ by 3 bp to remove the untemplated guanines added during cDNA synthesis. Libraries were sequenced with 50-60 million reads per sample.

References

Adachi, N. and M. R. Lieber (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell 109(7): 807-809.

Adachi, N. and M. R. Lieber (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell 109(7): 807-809.

Agger, K., et al. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 449(7163): 731-734.

Aggarwal B, and Calvi BR (2004). Chromatin regulates origin activity in Drosophila follicle cells. Nature 430(6997):372-6

Akkers, R. C., et al. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev Cell 17(3): 425-434.

Alabert, C., et al. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev 29(6): 585-590.

Alabert, C. and A. Groth (2012). Chromatin replication and epigenome maintenance. Nat Rev Mol Cell Biol 13(3): 153-167.

Aladjem, M. I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat Rev Genet 8(8): 588-600.

Albertson, D. G. and J. N. Thomson (1982). The kinetochores of Caenorhabditis elegans. Chromosoma 86(3): 409-428.

Allen, M. A., et al. (2011). A global analysis of C. elegans trans-splicing. Genome Res 21(2): 255-264.

Allen, M. A., et al. (2011). A global analysis of C. elegans trans-splicing. Genome Res 21(2): 255-264.

Almer, A., et al. (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. Embo j 5(10): 2689-2696.

Almouzni, G., et al. (1994). Histone acetylation influences both gene expression and development of Xenopus laevis. Dev Biol 165(2): 654-669.

Almouzni, G., et al. (1990). Competition between transcription complex assembly and chromatin assembly on replicating DNA. Embo j 9(2): 573-582.

Ambros, V. and H. R. Horvitz (1987). The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. Genes Dev 1(4): 398-414.

Amodeo, A. A., et al. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proc Natl Acad Sci U S A 112(10): E1086-1095.

Anderson, S. and M. L. DePamphilis (1979). Metabolism of Okazaki fragments during simian virus 40 DNA replication. J Biol Chem 254(22): 11495-11504.

Aoki, F., et al. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Dev Biol 181(2): 296-307.

Aoki, K., et al. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. Embo j 26(24): 5007-5019.

Aparicio, O. M., et al. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A 96(16): 9130-9135.

Aparicio, O. M., et al. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A 96(16): 9130-9135.

Arico, J. K., et al. (2011). Epigenetic patterns maintained in early Caenorhabditis elegans embryos can be established by gene activity in the parental germ cells. PLoS Genet 7(6): e1001391.

Ashe, A., et al. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell 150(1): 88-99.

Avgousti, D. C., et al. (2012). CSR-1 RNAi pathway positively regulates histone expression in C. elegans. Embo j 31(19): 3821-3832.

Bae, S. H. and Y. S. Seo (2000). Characterization of the enzymatic properties of the yeast dna2 Helicase/endonuclease suggests a new model for Okazaki fragment processing. J Biol Chem 275(48): 38022-38031.

Bannister, A. J. and T. Kouzarides (2011). Regulation of chromatin by histone modifications. Cell Res 21(3): 381-395.

Bao, Z., et al. (2008). Control of cell cycle timing during C. elegans embryogenesis. Dev Biol 318(1): 65-72.

Barnes, T. M., et al. (1995). Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. Genetics 141(1): 159-179.

Barriere, A. and M. A. Felix (2005). High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. Curr Biol 15(13): 1176-1184.

Bastiani, C. A., et al. (2003). Caenorhabditis elegans Galphaq regulates egg-laying behavior via a PLCbeta-independent and serotonin-dependent signaling pathway and likely functions both in the nervous system and in muscle. Genetics 165(4): 1805-1822.

Batista, P. J., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell 31(1): 67-78.

Baugh, L. R. (2013). To grow or not to grow: nutritional control of development during Caenorhabditis elegans L1 arrest. Genetics 194(3): 539-555.

Baugh, L. R., et al. (2003). Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. Development 130(5): 889-900.

Bektesh, S. L. and D. I. Hirsh (1988). C. elegans mRNAs acquire a spliced leader through a trans-splicing mechanism. Nucleic Acids Res 16(12): 5692.

Bell, S. P. and B. Stillman (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 357(6374): 128-134.

Benayoun, B. A., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell 158(3): 673-688.

Bender, L. B., et al. (2006). MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. Development 133(19): 3907-3917.

Berman, J. R. and C. Kenyon (2006). Germ-cell loss extends C. elegans life span through regulation of DAF-16 by kri-1 and lipophilic-hormone signaling. Cell 124(5): 1055-1068.

Bernstein, B. E., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2): 315-326.

Billi, A. C., et al. (2014). Endogenous RNAi pathways in C. elegans. WormBook: 1-49.

Bird, A. (2007). Perceptions of epigenetics. Nature 447(7143): 396-398.

Blumenthal, A. B., et al. (1974). The units of DNA replication in Drosophila melanogaster chromosomes. Cold Spring Harb Symp Quant Biol 38: 205-223.

Blumenthal, T., et al. (2015). Operon and non-operon gene clusters in the C. elegans genome. WormBook: 1-20.

Blumenthal, T. and K. S. Gleason (2003). Caenorhabditis elegans operons: form and function. Nat Rev Genet 4(2): 112-120.

Blumenthal, T. and J. Spieth (1996). Gene structure and organization in Caenorhabditis elegans. Curr Opin Genet Dev 6(6): 692-698.

Blumenthal, T. and J. Thomas (1988). Cis and trans mRNA splicing in C. elegans. Trends Genet 4(11): 305-308.

Blythe, S. A. and E. F. Wieschaus (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. Elife 5.

Boeck, M. E., et al. (2016). The time-resolved transcriptome of C. elegans. Genome Res 26(10): 1441-1450.

Bonn, S., et al. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44(2): 148-156.

Boveri T. Ein geschlechtlich erzeugter Organismus ohne mütterliche Eigenschaften. Sitz Gesel Morph u Physiol Müchen 5, 73–83 Trans by T H Morgan 1893, as An organism produced sexually without characteristics of the mother Am Naturalist. 1889;27:222–232.

Bowerman, B., et al. (1993). The maternal gene skn-1 encodes a protein that is distributed unequally in early C. elegans embryos. Cell 74(3): 443-452.

Boyer, L. A., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441(7091): 349-353.

Braendle, C. and M. A. Felix (2008). Plasticity and errors of a robust developmental system in different environments. Dev Cell 15(5): 714-724.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics 77(1): 71-94.

Brent, R. and M. Ptashne (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43(3 Pt 2): 729-736.

Buckley, B. A., et al. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature 489(7416): 447-451.

Buratowski, S., et al. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56(4): 549-561.

Burkhart, K. B., et al. (2011). A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. PLoS Genet 7(8): e1002249.

Burton, N. O., et al. (2011). Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. Proc Natl Acad Sci U S A 108(49): 19683-19688.

Campbell, A. C. and D. L. Updike (2015). CSR-1 and P granules suppress sperm-specific transcription in the C. elegans germline. Development 142(10): 1745-1755.

Cao, R., et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298(5595): 1039-1043.

Capowski, E. E., et al. (1991). Identification of grandchildless loci whose products are required for normal germ-line development in the nematode Caenorhabditis elegans. Genetics 129(4): 1061-1072.

Carmell, M. A. and G. J. Hannon (2004). RNase III enzymes and the initiation of gene silencing. Nat Struct Mol Biol 11(3): 214-218.

Cayrou, C. et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. 21(9):1438-49.

Cecere, G., et al. (2014). Global effects of the CSR-1 RNA interference pathway on the transcriptional landscape. Nat Struct Mol Biol 21(4): 358-365.

Chekanova, J. A., et al. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell 131(7): 1340-1353.

Chen, K., et al. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2: e00861.

Chen, R. A., et al. (2013). The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. Genome Res 23(8): 1339-1347.

Chen, R. A., et al. (2014). Extreme HOT regions are CpG-dense promoters in C. elegans and humans. Genome Res 24(7): 1138-1146.

Chrivia, J. C., et al. (1993). Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature 365(6449): 855-859.

Chuang, J. H. and H. Li (2004). Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome. PLoS Biol 2(2): E29.

Chuang, P. T., et al. (1994). DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome. Cell 79(3): 459-474.

Claycomb, J. M., et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell 139(1): 123-134.

Coleman, R. T. and G. Struhl (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science 356(6333).

Collart, C., et al. (2013). Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science 341(6148): 893-896.

Collart, C., et al. (2017). Chk1 Inhibition of the Replication Factor Drf1 Guarantees Cell-Cycle Elongation at the Xenopus laevis Mid-blastula Transition. Dev Cell 42(1): 82-96.e83.

Conine, C. C., et al. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proc Natl Acad Sci U S A 107(8): 3588-3593.

Conine, C. C., et al. (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. elegans. Cell 155(7): 1532-1544.

Conte, D., Jr. and C. C. Mello (2010). Primal RNAs: The end of the beginning? Cell 140(4): 452-454.

Cook, D. E., et al. (2017). CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res 45(D1): D650-d657.

Core, L. J., et al. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 46(12): 1311-1320.

Core, L. J., et al. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322(5909): 1845-1848.

Coulson, A., et al. (1986). Toward a physical map of the genome of the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A 83(20): 7821-7825.

Crane, E., et al. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523(7559): 240-244.

Crittenden, S. L., et al. (1994). GLP-1 is localized to the mitotic region of the C. elegans germ line. Development 120(10): 2901-2911.

Cui M., et al. (2008). Genes involved in pre-mRNA 3' end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen. PNAS 105(43): 1665-16670.

Cutter, A. D., et al. (2009). Evolution of the Caenorhabditis elegans genome. Mol Biol Evol 26(6): 1199-1234.

Danis, E., et al. (2004). Specification of a DNA replication origin by a transcription complex. Nat Cell Biol 6(8): 721-730.

Das, P. P., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell 31(1): 79-90.

Daugherty, A. C., et al. (2017). Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. Genome Res 27(12): 2096-2107.

de Albuquerque, B. F., et al. (2015). Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans. Dev Cell 34(4): 448-456.

De Santa, F., et al. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8(5): e1000384.

Dekker, J., et al. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14(6): 390-403.

Dekker, J. and T. Misteli (2015). Long-Range Chromatin Interactions. Cold Spring Harb Perspect Biol 7(10): a019356.

Dekker, J., et al. (2002). Capturing chromosome conformation. Science 295(5558): 1306-1311.

Devbhandari, S., et al. (2017). Chromatin Constrains the Initiation and Elongation of DNA Replication. Mol Cell 65(1): 131-141.

Dijkwel, P. A., et al. (1991). Mapping of replication initiation sites in mammalian genomes by two-dimensional gel analysis: stabilization and enrichment of replication intermediates by isolation on the nuclear matrix. Mol Cell Biol 11(8): 3850-3859.

Dileep, V., et al. (2015). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res 25(8): 1104-1113.

Dimitrova, D. S. and D. M. Gilbert (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Mol Cell 4(6): 983-993.

Dixon, J. R., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398): 376-380.

Draper, B. W., et al. (1996). MEX-3 is a KH domain protein that regulates blastomere identity in early C. elegans embryos. Cell 87(2): 205-216.

Duchaine, T. F., et al. (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. Cell 124(2): 343-354.

Dynan, W. S. and R. Tjian (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35(1): 79-87.

Dynlacht, B. D., et al. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell 66(3): 563-576.

Eckmann, C. R., et al. (2002). GLD-3, a bicaudal-C homolog that inhibits FBF to control germline sex determination in C. elegans. Dev Cell 3(5): 697-710.

Edgar, B. A., et al. (1986). Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell 44(2): 365-372.

Edgar, B. A. and G. Schubiger (1986). Parameters controlling transcriptional activation during early Drosophila development. Cell 44(6): 871-877.

Edgar, L. G. and J. D. McGhee (1988). DNA synthesis and the control of embryonic gene expression in C. elegans. Cell 53(4): 589-599.

Edgar, L. G., et al. (1994). Early transcription in Caenorhabditis elegans embryos. Development 120(2): 443-451.

Eissenberg, J. C. and A. Shilatifard (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol 339(2): 240-249.

Elewa, A., et al. (2015). POS-1 Promotes Endo-mesoderm Development by Inhibiting the Cytoplasmic Polyadenylation of neg-1 mRNA. Dev Cell 34(1): 108-118.

Elling, A. A., et al. (2007). Divergent evolution of arrested development in the dauer stage of Caenorhabditis elegans and the infective stage of Heterodera glycines. Genome Biol 8(10): R211.

Erdelyi, P., et al. (2017). A Network of Chromatin Factors Is Regulating the Transition to Postembryonic Development in Caenorhabditis elegans. G3 (Bethesda) 7(2): 343-353.

Erzberger, J. P. and J. M. Berger (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct 35: 93-114.

Euling, S. and V. Ambros (1996). Heterochronic genes control cell cycle progress and developmental competence of C. elegans vulva precursor cells. Cell 84(5): 667-676.

Evans, D., et al. (2001). A complex containing CstF-64 and the SL2 snRNP connects mRNA 3' end formation and trans-splicing in C. elegans operons. Genes Dev 15(19): 2562-2571.

Evans, K. J., et al. (2016). Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. Proc Natl Acad Sci U S A 113(45): E7020-e7029.

Evans, K. J., et al. (2016). Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. Proc Natl Acad Sci U S A 113(45): E7020-e7029.

Farrell, J. A. and P. H. O'Farrell (2013). Mechanism and regulation of Cdc25/Twine protein destruction in embryonic cell-cycle remodeling. Curr Biol 23(2): 118-126.

Fassnacht, C., et al. (2018). The CSR-1 endogenous RNAi pathway ensures accurate transcriptional reprogramming during the oocyte-to-embryo transition in Caenorhabditis elegans. PLoS Genet 14(3): e1007252.

Felix, M. A. and C. Braendle (2010). The natural history of Caenorhabditis elegans. Curr Biol 20(22): R965-969.

Felix, M. A. and C. Braendle (2010). The natural history of Caenorhabditis elegans. Curr Biol 20(22): R965-969.

Fennessy, R. T. and T. Owen-Hughes (2016). Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. Nucleic Acids Res 44(15): 7189-7203.

Ferguson, K. C., et al. (1996). The SL1 trans-spliced leader RNA performs an essential embryonic function in Caenorhabditis elegans that can also be supplied by SL2 RNA. Genes Dev 10(12): 1543-1556.

Ferguson, K. C., et al. (1996). The SL1 trans-spliced leader RNA performs an essential embryonic function in Caenorhabditis elegans that can also be supplied by SL2 RNA. Genes Dev 10(12): 1543-1556.

Fire, A., et al. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development 113(2): 503-514.

Fire, A., et al. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391(6669): 806-811.

Fire, A. Z. (2007). Gene silencing by double-stranded RNA. Cell Death Differ 14(12): 1998-2012.

Fischer, S. E., et al. (2013). Multiple small RNA pathways regulate the silencing of repeated and foreign genes in C. elegans. Genes Dev 27(24): 2678-2695.

Foe, V. E. and B. M. Alberts (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J Cell Sci 61: 31-70.

Frezal, L. and M. A. Felix (2015). C. elegans outside the Petri dish. Elife 4 ;4:e05849
.

Fuda, N. J., et al. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461(7261): 186-192.

Fuda, N. J., et al. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461(7261): 186-192.

Furuhashi, H., et al. (2010). Trans-generational epigenetic regulation of C. elegans primordial germ cells. Epigenetics Chromatin 3(1): 15.

Gan, H., et al. (2018). The Mcm2-Ctf4-Polalpha Axis Facilitates Parental Histone H3-H4 Transfer to Lagging Strands. Mol Cell 72(1): 140-151.e143.

Gassmann, R., et al. (2012). An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. Nature 484(7395): 534-537.

Gaydos, L. J., et al. (2012). Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. Cell Rep 2(5): 1169-1177.

Gaydos, L. J., et al. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science 345(6203): 1515-1518.

Georgescu, R., et al. (2017). Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. Proc Natl Acad Sci U S A 114(5): E697-e706.

Gerisch, B., et al. (2001). A hormonal signaling pathway influencing C. elegans metabolism, reproductive development, and life span. Dev Cell 1(6): 841-851.

Gerson-Gurwitz, A., et al. (2016). A Small RNA-Catalytic Argonaute Pathway Tunes Germline Transcript Levels to Ensure Embryonic Divisions. Cell 165(2): 396-409.

Gerstein, M. B., et al. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330(6012): 1775-1787.

Gerster, T., et al. (1986). During B-cell differentiation enhancer activity and transcription rate of immunoglobulin heavy chain genes are high before mRNA accumulation. Cell 45(1): 45-52.

Gibert M. A., Starck J., Beguet B. (1984). Role of the gonad cytoplasmic core during oogenesis of the nematode *Caenorhabditis elegans*. Biol. Cell **50**: 77–85.

Gilbert, D. M. (2002). Replication timing and transcriptional control: beyond cause and effect. Curr Opin Cell Biol 14(3): 377-383.

Gilbert, D. M., et al. (2010). Space and time in the nucleus: developmental control of replication timing and chromosome architecture. Cold Spring Harb Symp Quant Biol 75: 143-153.

Goodrich, J. A. and R. Tjian (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell 77(1): 145-156.

Goren, A., et al. (2008). DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. Genes Dev 22(10): 1319-1324.

Grant, P. A., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev 11(13): 1640-1650.

Greer, E. L., et al. (2014). A histone methylation network regulates transgenerational epigenetic memory in C. elegans. Cell Rep 7(1): 113-126.

Gregory, P. D., et al. (1998). Absence of Gcn5 HAT activity defines a novel state in the opening of chromatin at the PHO5 promoter in yeast. Mol Cell 1(4): 495-505.

Grewal, S. I. and S. C. Elgin (2007). Transcription and RNA interference in the formation of heterochromatin. Nature 447(7143): 399-406.

Grishok, A., et al. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 106(1): 23-34.

Groth, A., et al. (2007). Chromatin challenges during DNA replication and repair. Cell 128(4): 721-733.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. Nature 389(6649): 349-352.

Gu, S. G. and A. Fire (2010). Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma 119(1): 73-87.

Gu, W., et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell 36(2): 231-244.

Gu, W., et al. (2012). CapSeq and CIP-TAP map 5′ ends of Pol II transcripts and reveal capped-small RNAs as *C. elegans* piRNA precursors. Cell 151(7): 1488-1500.

Guang, S., et al. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature 465(7301): 1097-1101.

Guang, S., et al. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science 321(5888): 537-541.

Guiliano, D. B. and M. L. Blaxter (2006). Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genet 2(11): e198.

Gullerova, M. and N. J. Proudfoot (2012). Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. Nat Struct Mol Biol 19(11): 1193-1201.

Guo, S. and K. J. Kemphues (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81(4): 611-620.

Hamatani T. et al. (2004). Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell (6): 117-131.

Han, T., et al. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc Natl Acad Sci U S A 106(44): 18674-18679.

Harvey, S. A., et al. (2013). Identification of the zebrafish maternal and paternal transcriptomes. Development 140(13): 2703-2710.

Hashimshony, T., et al. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519(7542): 219-222.

Hashimshony, T., et al. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519(7542): 219-222.

Hashimshony, T., et al. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2(3): 666-673.

Hassan, A. H., et al. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell 111(3): 369-379.

Hatzis, P. and I. Talianidis (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell 10(6): 1467-1477.

Hebeisen, M., and Roy, R. (2008). CDC-25.1 stability is regulated by distinct domains to restrict cell division during embryogenesis. Development 135(7): 1259-1269.

Heller, R. C., et al. (2011). Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. Cell 146(1): 80-91.

Henderson, I. R. and S. E. Jacobsen (2007). Epigenetic inheritance in plants. Nature 447(7143): 418-424.

Henderson, S. T., et al. (1994). lag-2 may encode a signaling ligand for the GLP-1 and LIN-12 receptors of C. elegans. Development 120(10): 2913-2924.

Henriques, T., et al. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev 32(1): 26-41.

Heyn, P., et al. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep 6(2): 285-292.

Hirsh, D., et al. (1976). Development of the reproductive system of Caenorhabditis elegans. Dev Biol 49(1): 200-219.

Ho, C. K. and S. Shuman (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell 3(3): 405-411.

Ho, J. W., et al. (2014). Comparative analysis of metazoan chromatin organization. Nature 512(7515): 449-452.

Hodgkin, J. (1986). Sex determination in the nematode C. elegans: analysis of tra-3 suppressors and characterization of fem genes. Genetics 114(1): 15-52.

Holdeman, R., et al. (1998). MES-2, a maternal protein essential for viability of the germline in Caenorhabditis elegans, is homologous to a Drosophila Polycomb group protein. Development 125(13): 2457-2467.

Holstege, F. C., et al. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. Embo j 15(7): 1666-1677.

Hope, I. A. and K. Struhl (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46(6): 885-894.

Hsin, H. and C. Kenyon (1999). Signals from the reproductive system regulate the lifespan of C. elegans. Nature 399(6734): 362-366.

Huang, Y., et al. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science 312(5774): 748-751.

Hug, C. B., et al. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell 169(2): 216-228.e219.

Hutter, H. and R. Schnabel (1994). glp-1 and inductions establishing embryonic axes in C. elegans. Development 120(7): 2051-2064.

Hyrien, O. et al. (1995). Transition in specification of embryonic metazoan DNA replication origins. Science 270(5238):994-7

Iizuka M, and Stillman B. (1999). Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J Biol Chem, 274(33):23027-34.

Jacob, F., et al. (1960). [Operon: a group of genes with the expression coordinated by an operator]. C R Hebd Seances Acad Sci 250: 1727-1729.

Jacob, Y., et al. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol 16(7): 763-768.

Jenuwein, T. and C. D. Allis (2001). Translating the histone code. Science 293(5532): 1074-1080.

Jiao, L. and X. Liu (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science 350(6258): aac4383.

Johnson, A. and M. O'Donnell (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74: 283-315.

Jones, R. S. and W. M. Gelbart (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol 13(10): 6357-6366.

Jonkers, I., et al. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife 3: e02407.

Kagey, M. H., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314): 430-435.

Kaikkonen, M. U., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell 51(3): 310-325.

Kamath, R. S., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421(6920): 231-237.

Kane, D. A. and C. B. Kimmel (1993). The zebrafish midblastula transition. Development 119(2): 447-456.

Katz, D. J., et al. (2009). A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. Cell 137(2): 308-320.

Ke, Y., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell 170(2): 367-381.e320.

Kelly, W. G. (2014). Transgenerational epigenetics in the germline cycle of Caenorhabditis elegans. Epigenetics Chromatin 7(1): 6.

Kelly, W. G. and A. Fire (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development 125(13): 2451-2456.

Kelly, W. G., et al. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics 146(1): 227-238.

Kenyon, C., et al. (1993). A C. elegans mutant that lives twice as long as wild type. Nature 366(6454): 461-464.

Ketting, R. F., et al. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev 15(20): 2654-2659.

Kim, T. and S. Buratowski (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. Cell 137(2): 259-272.

Kim, Y. J., et al. (2018). Global transcriptional activity dynamics reveal functional enhancer RNAs. Genome Res.

Klosin, A., et al. (2017). Transgenerational transmission of environmental information in C. elegans. Science 356(6335): 320-323.

Koc, A., et al. (2004). Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. J Biol Chem 279(1): 223-230.

Korf, I., et al. (1998). The Polycomb group in Caenorhabditis elegans and maternal control of germline development. Development 125(13): 2469-2478.

Kornberg, R. D. and J. O. Thomas (1974). Chromatin structure; oligomers of the histones. Science 184(4139): 865-868.

Kriegstein, H. J. and D. S. Hogness (1974). Mechanism of DNA replication in Drosophila chromosomes: structure of replication forks and evidence for bidirectionality. Proc Natl Acad Sci U S A 71(1): 135-139.

Krogan, N. J., et al. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol 23(12): 4207-4218.

Kruesi, W. S., et al. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. Elife 2: e00808.

Krumm, A., et al. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev 9(5): 559-572.

Kuersten, S., et al. (1997). Relationship between 3' end formation and SL2-specific trans-splicing in polycistronic Caenorhabditis elegans pre-mRNA processing. Rna 3(3): 269-278.

Kuhn, C. D. and L. Joshua-Tor (2013). Eukaryotic Argonautes come into focus. Trends Biochem Sci 38(5): 263-271.

Kurat, C. F., et al. (2017). Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. Mol Cell 65(1): 117-130.

Kuzin, B., et al. (1994). The Drosophila trithorax gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. Genes Dev 8(20): 2478-2490.

Larsen, P. L., et al. (1995). Genes that regulate both development and longevity in Caenorhabditis elegans. Genetics 139(4): 1567-1583.

Laubichler, M. D. and E. H. Davidson (2008). Boveri's long experiment: Sea urchin merogones and the establishment of the role of nuclear chromosomes in development. Dev Biol 314(1): 1-11.

Lawlis, S. J., et al. (1996). Chromosome architecture can dictate site-specific initiation of DNA replication in Xenopus egg extracts. J Cell Biol 135(5): 1207-1218.

Lawrence, J. (1999). Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr Opin Genet Dev 9(6): 642-648.

Lee, C. K., et al. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36(8): 900-905.

Lee, H. C., et al. (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell 150(1): 78-87.

Lee, L. A. and T. L. Orr-Weaver (2003). Regulation of cell cycles in Drosophila development: intrinsic and extrinsic cues. Annu Rev Genet 37: 545-578.

Lee, M. T., et al. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol 30: 581-613.

Lee, M. T., et al. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 503(7476): 360-364.

Leive, L. and V. Kollin (1967). Synthesis, utilization and degradation of lactose operon mRNA in Escherichia coli. J Mol Biol 24(2): 247-259.

Lerner, K. and P. Goldstein (1988). Electron microscopic autoradiographic analysis: evidence of RNA transcription along pachytene chromosomes of rad-4, him-4 and wild-type Caenorhabditis elegans. Cytobios 55(220): 51-61.

Levin, D. S., et al. (1997). An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. Proc Natl Acad Sci U S A 94(24): 12863-12868.

Levin, M., et al. (2016). The mid-developmental transition and the evolution of animal body plans. Nature 531(7596): 637-641.

Levin, M., et al. (2016). The mid-developmental transition and the evolution of animal body plans. Nature 531(7596): 637-641.

Levin, M., et al. (2012). Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev Cell 22(5): 1101-1108.

Levine, M. and J. L. Manley (1989). Transcriptional repression of eukaryotic promoters. Cell 59(3): 405-408.

Levine, M. and R. Tjian (2003). Transcription regulation and animal diversity. Nature 424(6945): 147-151.

Li, B., et al. (2007). The role of chromatin during transcription. Cell 128(4): 707-719.

Li, J. J. and T. J. Kelly (1984). Simian virus 40 DNA replication in vitro. Proc Natl Acad Sci U S A 81(22): 6973-6977.

Li, P. C., et al. (2013). Mutations disrupting histone methylation have different effects on replication timing in S. pombe centromere. PLoS One 8(5): e61464.

Li, W., et al. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet 17(4): 207-223.

Liang, H. L., et al. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456(7220): 400-403.

Lieberman-Aiden, E., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950): 289-293.

Lindeman, L. C., et al. (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Dev Cell 21(6): 993-1004.

Lindstrom, D. L. and G. A. Hartzog (2001). Genetic interactions of Spt4-Spt5 and TFIIS with the RNA polymerase II CTD and CTD modifying enzymes in Saccharomyces cerevisiae. Genetics 159(2): 487-497.

Liu, Y. et al. (2001). Interplay between AAUAAA and the trans-splice site in processing of a Caenorhabditis elegans operon pre-mRNA. RNA: 7(2): 176-181.

Liu, J., et al. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science 305(5689): 1437-1441.

Liu, K. S. and P. W. Sternberg (1995). Sensory regulation of male mating behavior in Caenorhabditis elegans. Neuron 14(1): 79-89.

Liu, T., et al. (2011). Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res 21(2): 227-236.

Long, H. K., et al. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167(5): 1170-1187.

Longman, D. et al. (2000). Functional characterization of SR and SR-related genes in *C. elegans*. EMBO 19(7): 1625-1637.

Lu, X., et al. (2009). Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development 136(12): 2101-2110.

Lubelsky Y, Prinz JA, DeNapoli L, Li Y, Belsky JA, MacAlpine DM
Genome Res. 2014 Jul; 24(7):1102-14.

Lucio-Eterovic, A. K., et al. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. Proc Natl Acad Sci U S A 107(39): 16952-16957.

Lue, N. F., et al. (1989). Initiation by yeast RNA polymerase II at the adenoviral major late promoter in vitro. Science 246(4930): 661-664.

Luger, K., et al. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389(6648): 251-260.

Malik, S., et al. (1991). Sequence of general transcription factor TFIIB and relationships to other initiation factors. Proc Natl Acad Sci U S A 88(21): 9553-9557.

Mandal, S. S., et al. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc Natl Acad Sci U S A 101(20): 7572-7577.

Mango, S. E., et al. (1994). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. Development 120(10): 3019-3031.

Maniar, J. M. and A. Z. Fire (2011). EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. Curr Biol 21(6): 449-459.

Mao, H., et al. (2015). The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. Curr Biol 25(18): 2398-2403.

Margueron, R., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461(7265): 762-767.

Matranga, C., et al. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123(4): 607-620.

Maxwell, C. S., et al. (2012). Nutritional control of mRNA isoform expression during developmental arrest and recovery in C. elegans. Genome Res 22(10): 1920-1929.

McGuffee, S. R., et al. (2013). Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. Mol Cell 50(1): 123-135.

Meister, G., et al. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15(2): 185-197.

Mello, C. C., et al. (1992). The pie-1 and mex-1 genes and maternal control of blastomere identity in early C. elegans embryos. Cell 70(1): 163-176.

Mello, C. C., et al. (1996). The PIE-1 protein and germline specification in C. elegans embryos. Nature 382(6593): 710-712.

Miller, T., et al. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A 98(23): 12902-12907.

Minkina, O. and C. P. Hunter (2017). Stable Heritable Germline Silencing Directs Somatic Silencing at an Endogenous Locus. Mol Cell 65(4): 659-670 e655.

Mizuguchi, G., et al. (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. Mol Cell 1(1): 141-150.

Moazed, D. (2011). Mechanisms for the inheritance of chromatin states. Cell 146(4): 510-518.

Montgomery, M. K. and A. Fire (1998). Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends Genet 14(7): 255-258.

Moorman, C., et al. (2006). Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc Natl Acad Sci U S A 103(32): 12027-12032.

Mujtaba, S., et al. (2007). Structure and acetyl-lysine recognition of the bromodomain. Oncogene 26(37): 5521-5527.

Murakami, K. S., et al. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science 296(5571): 1285-1290.

Muramoto, T., et al. (2010). Methylation of H3K4 Is required for inheritance of active transcriptional states. Curr Biol 20(5): 397-406.

Narendra, V., et al. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev 30(24): 2657-2662.

Neil, H., et al. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457(7232): 1038-1042.

Newport, J. and M. Kirschner (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 30(3): 675-686.

Newport, J. and M. Kirschner (1982). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell 30(3): 687-696.

Nichols, M. H. and V. G. Corces (2015). A CTCF Code for 3D Genome Architecture. Cell 162(4): 703-705.

Nieduszynski, C. A., et al. (2007). OriDB: a DNA replication origin database. Nucleic Acids Res 35(Database issue): D40-46.

Noble, D. (2015). Conrad Waddington and the origin of epigenetics. J Exp Biol 218(Pt 6): 816-818.
O'Donnell, M., et al. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harb Perspect Biol 5(7).

Ogryzko, V. V., et al. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell 87(5): 953-959.
suggest that p300/CBP acetylates nucleosomes in concert with PCAF.

Okamura, K., et al. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol 15(9): 998.

Orkin, S. H. and K. Hochedlinger (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell 145(6): 835-850.

Orlando, D. A., et al. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep 9(3): 1163-1170.

Ortiz, M. A., et al. (2014). A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans. G3 (Bethesda) 4(9): 1765-1772.

Pak, J. and A. Fire (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science 315(5809): 241-244.

Parkinson, J., et al. (2004). A transcriptomic analysis of the phylum Nematoda. Nat Genet 36(12): 1259-1267.

Parrish, S. and A. Fire (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. Rna 7(10): 1397-1402.

Patel, T., et al. (2012). Removal of Polycomb repressive complex 2 makes C. elegans germ cells susceptible to direct conversion into specific somatic cell types. Cell Rep 2(5): 1178-1186.

Paulsen, J. E., et al. (1995). Phenotypic and molecular analysis of mes-3, a maternal-effect gene required for proliferation and viability of the germ line in C. elegans. Genetics 141(4): 1383-1398.

Perino, M. and G. J. Veenstra (2016). Chromatin Control of Developmental Dynamics and Plasticity. Dev Cell 38(6): 610-620.

Peters, A. H., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell 12(6): 1577-1589.

Petryk, N., et al. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. Science 361(6409): 1389-1392.

Petryk, N., et al. (2016). Replication landscape of the human genome. Nat Commun 7: 10208.

Pham, J. W., et al. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. Cell 117(1): 83-94.

Phatnani, H. P. and A. L. Greenleaf (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev 20(21): 2922-2936.

Phillips, C. M., et al. (2012). MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. Genes Dev 26(13): 1433-1444.

Phillips-Cremins, J. E., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153(6): 1281-1295.

Pilgrim, D. (1993). The genetic and RFLP characterization of the left end of linkage group III in Caenorhabditis elegans. Genome 36(4): 712-724.

Pokholok, D. K., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122(4): 517-527.

Pope, B. D., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature 515(7527): 402-405.

Potok, M. E., et al. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. Cell 153(4): 759-772.

Potter, P. C., et al. (1995). Immediate and delayed contact hypersensitivity to verbena plants. Contact Dermatitis 33(5): 343-346.

Pourkarimi, E., et al. (2016). Spatiotemporal coupling and decoupling of gene transcription with DNA replication origins during embryogenesis in C. elegans. Elife 5.

Powell-Coffman, J. A., et al. (1996). Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. Dev Biol 178(2): 472-483.

Prioleau, M. N., et al. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell 77(3): 439-449.

Pritchard, D. K. and G. Schubiger (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev 10(9): 1131-1142.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol 16(3): 272-278.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322(6081): 697-701.

Ptashne, M. and A. Gann (1997). Transcriptional activation by recruitment. Nature 386(6625): 569-577.

Qiu, H., et al. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Mol Cell 33(6): 752-762.

Raghuraman, M. K., et al. (2001). Replication dynamics of the yeast genome. Science 294(5540): 115-121.

Ramachandran, S. and S. Henikoff (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell 165(3): 580-592.

Rao, S. S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7): 1665-1680.

Rasmussen, E. B. and J. T. Lis (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A 90(17): 7923-7927.

Reijns, M. A. M., et al. (2015). Lagging-strand replication shapes the mutational landscape of the genome. Nature 518(7540): 502-506.

Reinhart, B. J. and D. P. Bartel (2002). Small RNAs correspond to centromere heterochromatic repeats. Science 297(5588): 1831.

Reinke, V. and A. D. Cutter (2009). Germline expression influences operon organization in the Caenorhabditis elegans genome. Genetics 181(4): 1219-1228. specific tissue can have profound effects on the evolution of genome organization.

Reinke, V., et al. (2004). Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. Development 131(2): 311-323.

Remus, D., et al. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell 139(4): 719-730.

Reveron-Gomez, N., et al. (2018). Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. Mol Cell.

Richter, J. D. and P. Lasko (2011). Translational control in oocyte development. Cold Spring Harb Perspect Biol 3(9): a002758.

Rivera-Mulia, J. C. and D. M. Gilbert (2016). Replicating Large Genomes: Divide and Conquer. Mol Cell 62(5): 756-765.

Rivin, C. J. and W. L. Fangman (1980). Replication fork rate and origin activation during the S phase of Saccharomyces cerevisiae. J Cell Biol 85(1): 108-115.

Robertson, S. and R. Lin (2015). The Maternal-to-Zygotic Transition in C. elegans. Curr Top Dev Biol 113: 1-42.

Robinson, M. D. and A. Oshlack (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11(3): R25.

Rocheleau, C. E., et al. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell 90(4): 707-716.

Rothe, M., et al. (1992). Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature 359(6391): 156-159.

Ryba, T., et al. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res 20(6): 761-770.

Saito, T. L., et al. (2013). The transcription start site landscape of C. elegans. Genome Res 23(8): 1348-1361.

Sarkies, P., et al. (2013). Competition between virus-derived and endogenous small RNAs regulates gene expression in Caenorhabditis elegans. Genome Res 23(8): 1258-1270.

Schauer, I. E. and W. B. Wood (1990). Early C. elegans embryos are transcriptionally active. Development 110(4): 1303-1317.

Schierenberg, E., et al. (1980). Cell lineages and developmental defects of temperature-sensitive embryonic arrest mutants in Caenorhabditis elegans. Dev Biol 76(1): 141-159.

Schmitges, F. W., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell 42(3): 330-341.

Schmitges, F. W., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell 42(3): 330-341.

Schnabel, R. and J. R. Priess (1997). Specification of Cell Fates in the Early Embryo. C. elegans II. nd, D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.

Schwacha, A. and S. P. Bell (2001). Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. Mol Cell 8(5): 1093-1104.

Scruggs, B. S., et al. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell 58(6): 1101-1112.

Segal, E., et al. (2006). A genomic code for nucleosome positioning. Nature 442(7104): 772-778.

Seth, M., et al. (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Dev Cell 27(6): 656-663.

Sexton, T., et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148(3): 458-472.

Seydoux, G. and R. E. Braun (2006). Pathway to totipotency: lessons from germ cells. Cell 127(5): 891-904.

Seydoux, G. and A. Fire (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in Caenorhabditis elegans. Development 120(10): 2823-2834.

Seydoux, G., et al. (1996). Repression of gene expression in the embryonic germ lineage of C. elegans. Nature 382(6593): 713-716.

Shakes, D. C. and S. Ward (1989). Initiation of spermiogenesis in C. elegans: a pharmacological and genetic analysis. Dev Biol 134(1): 189-200.

Shen, E. Z., et al. (2018). Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. Cell 172(5): 937-951.e918.

Shermoen, A. W. and P. H. O'Farrell (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell 67(2): 303-310.

Shi, Y. and C. Mello (1998). A CBP/p300 homolog specifies multiple differentiation pathways in Caenorhabditis elegans. Genes Dev 12(7): 943-955.

Shibahara, K. and B. Stillman (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell 96(4): 575-585.

Shibahara, K. and B. Stillman (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell 96(4): 575-585.

Shin, T. H. and C. C. Mello (2003). Chromatin regulation during C. elegans germline development. Curr Opin Genet Dev 13(5): 455-462.

Shirayama, M., et al. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell 150(1): 65-77.

Shirayama, M., et al. (2014). The Vasa Homolog RDE-12 engages target mRNA and multiple argonaute proteins to promote RNAi in C. elegans. Curr Biol 24(8): 845-851.

Shlyueva, D., et al. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15(4): 272-286.

Sigova, A. A., et al. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science 350(6263): 978-981.

Sijen, T., et al. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107(4): 465-476.

Simon, J. A. and R. E. Kingston (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10(10): 697-708.

Smardon, A., et al. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol 10(4): 169-178.

Smith, D. J. and I. Whitehouse (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483(7390): 434-438.

Sommer, R. J., et al. (1994). The evolution of cell lineage in nematodes. Dev Suppl: 85-95.

Song, J. J., et al. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305(5689): 1434-1437.

Soshnikova, N. and D. Duboule (2009). Epigenetic temporal control of mouse Hox genes in vivo. Science 324(5932): 1320-1323.

Spieth, J. and D. Lawson (2006). Overview of gene structure. WormBook: 1-10.

Spieth, J., et al. (2014). Overview of gene structure in C. elegans. WormBook: 1-18.

Stefani, G. and F. J. Slack (2008). Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol 9(3): 219-230.

Steiner, F. A. and S. Henikoff (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife 3: e02025.

Stillman, B. (2005). Origin recognition and the chromosome cycle. FEBS Lett 579(4): 877-884.

Stillman, B. (2005). Origin recognition and the chromosome cycle. FEBS Lett 579(4): 877-884.

Stoeckius, M., et al. (2014). Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. Embo j 33(16): 1751-1766.

Strome, S. (1986). Fluorescence visualization of the distribution of microfilaments in gonads and early embryos of the nematode Caenorhabditis elegans. J Cell Biol 103(6 Pt 1): 2241-2252.

Strome, S., et al. (1995). Transformation of the germ line into muscle in mes-1 mutant embryos of C. elegans. Development 121(9): 2961-2972.

Strome, S. and W. B. Wood (1982). Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of Caenorhabditis elegans. Proc Natl Acad Sci U S A 79(5): 1558-1562.

Sulston, J. E. and H. R. Horvitz (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol 56(1): 110-156.

Sulston, J. E., et al. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 100(1): 64-119.

Sutton, R. E. and J. C. Boothroyd (1986). Evidence for trans splicing in trypanosomes. Cell 47(4): 527-535.

Swigut, T. and J. Wysocka (2007). H3K27 demethylases, at long last. Cell 131(1): 29-32.

Tabara, H., et al. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99(2): 123-132.

Tam, O. H., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453(7194): 534-538.

Tanaka, S. and H. Araki (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. Cold Spring Harb Perspect Biol 5(12): a010371.

Thomas, J. D., et al. (1988). The C. elegans trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell 54(4): 533-539.

Tijsterman, M., et al. (2002). The genetics of RNA silencing. Annu Rev Genet 36: 489-519.

Tsai, H. Y., et al. (2015). A ribonuclease coordinates siRNA amplification and mRNA cleavage during RNAi. Cell 160(3): 407-419.

Tsukiyama, T. and C. Wu (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell 83(6): 1011-1020.

Tsurimoto, T. and B. Stillman (1991). Replication factors required for SV40 DNA replication in vitro. II. Switching of DNA polymerase alpha and delta during initiation of leading and lagging strand synthesis. J Biol Chem 266(3): 1961-1968.

Tu, S., et al. (2015). Comparative functional characterization of the CSR-1 22G-RNA pathway in Caenorhabditis nematodes. Nucleic Acids Res 43(1): 208-224.

Tursun, B., et al. (2011). Direct conversion of C. elegans germ cells into specific neuron types. Science 331(6015): 304-308.

Tzur, Y. B., et al. (2018). Spatiotemporal Gene Expression Analysis of the Caenorhabditis elegans Germline Uncovers a Syncytial Expression Switch. Genetics 210(2): 587-605.

Uhlmann, F. and K. Nasmyth (1998). Cohesion between sister chromatids must be established during DNA replication. Curr Biol 8(20): 1095-1101.

Ulianov, S. V., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res 26(1): 70-84.

Unhavaithaya, Y., et al. (2002). MEP-1 and a homolog of the NURD complex component Mi-2 act together to maintain germline-soma distinctions in C. elegans. Cell 111(7): 991-1002.

Updike, D. L. and S. Strome (2009). A genomewide RNAi screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans. Genetics 183(4): 1397-1419.

Vasale, J. J., et al. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. Proc Natl Acad Sci U S A 107(8): 3582-3587.

Vastenhouw, N. L., et al. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature 464(7290): 922-926.

Veenstra, G. J., et al. (1999). Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition. Mol Cell Biol 19(12): 7972-7982.

Veenstra, G. J., et al. (2000). Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus. Science 290(5500): 2312-2315.

Venkatesh, S., et al. (2016). Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat Commun 7: 13610.

Verrijzer, C. P., et al. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. Cell 81(7): 1115-1125.

Victor, M., et al. (2002). HAT activity is essential for CBP-1-dependent transcription and differentiation in Caenorhabditis elegans. EMBO Rep 3(1): 50-55.

Vietri Rudan, M., et al. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep 10(8): 1297-1309.

Vietri Rudan, M. and S. Hadjur (2015). Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. Trends Genet 31(11): 651-660.

Villeneuve, A. M. and B. J. Meyer (1987). sdc-1: a link between sex determination and dosage compensation in C. elegans. Cell 48(1): 25-37.

Vogelauer, M., et al. (2002). Histone acetylation regulates the time of replication origin firing. Mol Cell 10(5): 1223-1233.

Waddington, C.H. (1956). The genetic assimilation of the bithorax phenotype. Evolution. 10: 1-3.

Waddington, C.H. (1957). *The strategy of the genes*. London: Allen and Unwin. Reprinted 2014.

Waga, S. and B. Stillman (1998). The DNA replication fork in eukaryotic cells. Annu Rev Biochem 67: 721-751.

Walser, C. B. and H. D. Lipshitz (2011). Transcript clearance during the maternal-to-zygotic transition. Curr Opin Genet Dev 21(4): 431-443.

Wang, D., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474(7351): 390-394.

Wang, D., et al. (2005). Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. Nature 436(7050): 593-597.

Ward, S. and J. S. Carrel (1979). Fertilization and sperm competition in the nematode Caenorhabditis elegans. Dev Biol 73(2): 304-321.

Warf, M. B., et al. (2012). Effects of ADARs on small RNA processing pathways in C. elegans. Genome Res 22(8): 1488-1498.

Waring, D. A. and C. Kenyon (1990). Selective silencing of cell communication influences anteroposterior pattern formation in C. elegans. Cell 60(1): 123-131.

Watanabe, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453(7194): 539-543.

Watson, J.D. and Crick, F.H. (1953). Genetical implications of the structure of deoxyribonucleic acid. Nature 171: 964-967.

Weber, F. and W. Schaffner (1985). Enhancer activity correlates with the oncogenic potential of avian retroviruses. Embo j 4(4): 949-956.

Wedeles, C. J., et al. (2013). Protection of germline gene expression by the C. elegans Argonaute CSR-1. Dev Cell 27(6): 664-671.

Weeks, J. R., et al. (1993). Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. Genes Dev 7(12a): 2329-2344.

White, J. G., et al. (1976). The structure of the ventral nerve cord of Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci 275(938): 327-348.

Wolffe, A. P. and D. Pruss (1996). Targeting chromatin disruption: Transcription regulators that acetylate histones. Cell 84(6): 817-819.

Wood, A., et al. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem 278(37): 34739-34742.

Wood, A., et al. (2007). Ctk complex-mediated regulation of histone methylation by COMPASS. Mol Cell Biol 27(2): 709-720.

Wood, W. B. (1991). Evidence from reversal of handedness in C. elegans embryos for early cell interactions determining cell fates. Nature 349(6309): 536-538.

Wu, S. F., et al. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res 21(4): 578-589.

Wysocka, J., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442(7098): 86-90.

Xu, L., et al. (2001). The Caenorhabditis elegans maternal-effect sterile proteins, MES-2, MES-3, and MES-6, are associated in a complex in embryos. Proc Natl Acad Sci U S A 98(9): 5061-5066.

Xu, Z., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature 457(7232): 1033-1037.

Yadav, T. and I. Whitehouse (2016). Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. Cell Rep 15(4): 715-723.

Yao, N.Y. and O'Donnell, M. (2016). Evolution of replication machines. Crit Rev Biochem Mol Biol 51(3): 135-149.

Yasuda, G. K., et al. (1991). Temporal regulation of gene expression in the blastoderm Drosophila embryo. Genes Dev 5(10): 1800-1812.

Yehuda, Y., et al. (2018). Germline DNA replication timing shapes mammalian genome composition. Nucleic Acids Res 46(16): 8299-8310.

Yigit, E., et al. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 127(4): 747-757.

Yuan, G. C., et al. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309(5734): 626-630.

Yuan, W., et al. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem 286(10): 7983-7989.

Yuryev, A. and J. L. Corden (1996). Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. Genetics 143(2): 661-671.

Yuzyuk, T., et al. (2009). The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in C. elegans embryos. Dev Cell 16(5): 699-710.

Zalts, H. and I. Yanai (2017). Developmental constraints shape the evolution of the nematode mid-developmental transition. Nat Ecol Evol 1(5): 113.

Zaratiegui, M., et al. (2011). RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. Nature 479(7371): 135-138.

Zaslaver, A., et al. (2011). Metazoan operons accelerate recovery from growth-arrested states. Cell 145(6): 981-992.

Zechner, E. L., et al. (1992). Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. J Biol Chem 267(6): 4045-4053.

Zenk, F., et al. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science 357(6347): 212-216.

Zhang, C., et al. (2012). The Caenorhabditis elegans RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. Curr Biol 22(10): 881-890.

Zhang, Y., et al. (2014). Canonical nucleosome organization at promoters forms during genome activation. Genome Res 24(2): 260-266.